

Aboul Ella Hassanien · Abdelkrim Haqiq ·
Ahmad Taher Azar · KC Santosh ·
M. A. Jabbar · Adam Słowik ·
Parthasarathy Subashini *Editors*

The 3rd International Conference on Artificial Intelligence and Computer Vision (AICV2023), March 5–7, 2023

Lecture Notes on Data Engineering and Communications Technologies

164

Series Editor

Fatos Xhafa, *Technical University of Catalonia, Barcelona, Spain*

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

Indexed by SCOPUS, INSPEC, EI Compendex.

All books published in the series are submitted for consideration in Web of Science.

Aboul Ella Hassanien · Abdelkrim Haqiq ·
Ahmad Taher Azar · KC Santosh · M. A. Jabbar ·
Adam Słowik · Parthasarathy Subashini
Editors

The 3rd International Conference on Artificial Intelligence and Computer Vision (AICV2023), March 5–7, 2023

Editors

Aboul Ella Hassanien 
Faculty of Computer and AI
Cairo University
Giza, Egypt

Ahmad Taher Azar
College of Computer and Information
Sciences
Prince Sultan University
Riyadh, Saudi Arabia

M. A. Jabbar
Vardhaman College of Engineering
Hyderabad, Telangana, India

Parthasarathy Subashini
Department of Computer Science
Avinashilingam University for Women
Coimbatore, Tamil Nadu, India

Abdelkrim Haqiq
Faculty of Sciences and Techniques
Hassan 1st University
Settat, Morocco

KC Santosh 
Department of Computer Science
University of South Dakota
Vermillion, SD, USA

Adam Słowik
Department of Electronics and Computer
Science
Koszalin University of Technology
Koszalin, Poland

ISSN 2367-4512

ISSN 2367-4520 (electronic)

Lecture Notes on Data Engineering and Communications Technologies

ISBN 978-3-031-27761-0

ISBN 978-3-031-27762-7 (eBook)

<https://doi.org/10.1007/978-3-031-27762-7>

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The 3rd International Conference on Artificial Intelligence and Computer Vision (AICV2023), which took place at Hassan 1st University, Marrakesh, Morocco, from March 5 to 7, 2023, is an international conference covering research and development in artificial intelligence and computer vision. The 3rd edition of AICV2023 is organized by the Scientific Research Group in Egypt (SRGE) and the Computer, Networks, Mobility and Modeling Laboratory (IR2M), Hassan 1st University, Faculty of Sciences and Techniques, Settat, Morocco, and the Automated Systems & Soft Computing Lab (ASSCL), Prince Sultan University, Riyadh, Saudi Arabia. AICV'2023 is organized to provide an international forum that brings together those actively involved in the areas of interest and reports on up-to-the-minute innovations and developments to summarize the state of the art and exchange ideas and advances in all aspects of artificial intelligence and computer vision. The conference proceedings has nine major tracks:

- Part-1 Deep learning and intelligent applications
- Part-2 Machine learning and applications
- Part-3 Image processing and computer vision
- Part-4 Sentiment analysis and recommendation systems
- Part-5 Data sciences and business-based applications
- Part-6 Metaheuristic algorithms-based applications
- Part-7 Software-defined network and telecommunication

All submissions were reviewed by three reviewers on average, with no distinction between papers submitted for all conference tracks. We are convinced that the quality and diversity of the topics covered will satisfy both the attendees and the readers of these conference proceedings. We express our sincere thanks to the plenary speakers and the international program committee members for formulating a rich technical program. We want to extend our sincere appreciation for the outstanding work contributed over many months by the organizing committee: local organization chair and publicity chair. We also wish to express our gratitude to the SRGE members for their assistance. We want to emphasize that the success of AICV2023 would not have been possible without the support of many committed volunteers who generously contributed their time, expertise and resources toward making the conference an unqualified success.

Aboul Ella Hassanien
Abdelkrim Haqiq
Ahmad Taher Azar
KC Santosh
M. A. Jabbar
Adam Słowik
Parthasarathy Subashini

Organization

International Advisory Board

Adel Mohamed Alimi	National Engineering School of Sfax, Tunisia
Dabia Ahmed Abouâinainen	Abdulrahman Bin Faisal University, Saudi Arabia
Rawya Rizk	Port Said University, Egypt
Youssef F. Rashed	Cairo University, Egypt

General Chairs

Abdelkrim Haqiq	FST, Hassan 1st University, Settat, Morocco
Aboul Ella Hassanien	Faculty of Computer and AI, Cairo University, Egypt
Ahmad Taher Azar	College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia; Leader of Automated Systems and Soft Computing Lab (ASSCL), Prince Sultan University, Riyadh, Saudi Arabia; Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt

Technical Program Chairs

KC Santosh	Department of Computer Science, University of South Dakota, USA
Akhil Jabbar Meerja	Vardhaman College of Engineering, India
Adam Słowik	Department of Electronics and Computer Science, Koszalin University of Technology, Koszalin, Poland
Parthasarathy Subashini	Avinashilingam University for Women, Coimbatore, India
Vaclav Snasel	VSB-Technical University of Ostrava, Czech Republic

Track Chairs

Track 1: Machine and Deep learning

Asadullah Shaikh	Najran University, Najran, Saudi Arabia
Abdulkareem Alzahrani	Albaha University, Saudi Arabia
Umang Singh	Institute of Technology and Science, Ghaziabad, India
Lahouari Ghouti	Automated Systems and Soft Computing Lab, Prince Sultan University, Riyadh, Saudi Arabia

Track 2: Computational Intelligence

Adam Slowik	Koszalin University of Technology, Koszalin, Poland
Said El Kafhali	Faculty of Sciences and Techniques, Settati, Morocco
Mohamed Tounsi	Automated Systems and Soft Computing Lab, Prince Sultan University, Riyadh, Saudi Arabia

Track 3: Mining and Data Analysis

Mohamed Hanini	Faculty of Sciences and Techniques, Settati, Morocco
Suliman Mohamed Fati	Automated Systems and Soft Computing Lab, Prince Sultan University, Riyadh, Saudi Arabia

Track 4: Robotics and Automation

Ammar Al-Mhdawi	Edge Hill University, UK
Nizar Rokbani	University of Sousse, Tunisia

Track 5: Image Processing and Computer Vision

Houssam Halmaoui	ISMAL, Rabat, Morocco
Khattab M. Ali Alheeti	University of Anbar, Iraq

Publicity Chairs

Azza Ahmed Abdo Ali	Abdulrahman Bin Faisal University, Saudi Arabia
Brahim Ouhbi	National School of Arts and Crafts (ENSAM), Meknès, Morocco
Driss Bouzidi	ENSIAS, Mohammed V University, Rabat, Morocco
Driss El Ouadghiri	Faculty of Sciences, Meknès, Morocco
Essaid Sabir	ENSEM, Hassan II University, Casablanca, Morocco
Lavika Goel	Malaviya, National Institute of Technology (NIT), Jaipur, India
Mohamed Nemiche	Faculty of Sciences, Agadir, Morocco
Younma El Hissi	ENCG, Sultan Moulay Slimane University, Beni Mellal, Morocco
Yasir Javed	Automated Systems and Soft Computing Lab, Prince Sultan University, Riyadh, Saudi Arabia
Syed Umar Amin	Automated Systems and Soft Computing Lab, Prince Sultan University, Riyadh, Saudi Arabia
Shrooq Alsenan	Automated Systems and Soft Computing Lab, Prince Sultan University, Riyadh, Saudi Arabia
Zafar Iqbal Khan	Automated Systems and Soft Computing Lab, Prince Sultan University, Riyadh, Saudi Arabia

Local Arrangement Committee Chairs

Abdellah Zaaloul	FP, Ibn Zohr University, Agadir, Morocco
El Mehdi Kandoussi	INPT, Rabat, Morocco
Iman Elmir	I2S, Hassan 1st University, Settat, Morocco
Jaouad Dabounou	FST, Hassan 1st University, Settat, Morocco
Mohamed Hanini	FST, Hassan 1st University, Settat, Morocco
Mohamed Nabil	FS, Chouaïb Doukali University, El Jadida, Morocco
Said El Kafhali	FST, Hassan 1st University, Settat, Morocco

Technical Program Committee Members

Abdelghani Bentahar	FST, Hassan 1st University, Settat, Morocco
Abdelhay Haqiq	ESI, Mohammed V University, Rabat, Morocco
Abdellah Jamali	ENSA, Berchid, Morocco
Abdellah Ezzati	FST, Hassan 1st University, Settat, Morocco

Abdellah Najid	INPT, Rabat, Morocco
Abdellah Zaaloul	Ibn Zohr University, Agadir, Morocco
Abdellatif Kobbane	ENSIAS, Mohammed V University, Rabat, Morocco
Abdelmajid Hajami	FST, Hassan 1st University, Settat, Morocco
Abderrahim Benihssane	FS El Jadida, Morocco
Abderrahim Marzouk	FST, Hassan 1st University, Settat, Morocco
Abderrahim Sekkaki	FS Ain Chock, Hassan II University, Casablanca, Morocco
Ahmed Boujnoui	FST, Hassan 1st University, Settat, Morocco
Adam Slowik	Koszalin University of Technology, Koszalin, Poland
Adnane Founoun	FST, Hassan 1st University, Settat, Morocco
Adnane Latif	ENSA, Cadi Ayyad University, Marrakesh, Morocco
Ahlame Begdouri	FST, My Abdellah University, Fès, Morocco
Ahmed Zellou	ENSIAS, Mohammed V University, Rabat, Morocco
Ajith Abraham	MIRLabs, Washington, USA
Akhil Jabbar Meerja	Vardhaman College of Engineering, India
Alaa M. Khamis	General Motors Canada, Canada
Amine Ben Makhlof	FST, Hassan 1st University, Settat, Morocco
Amine Berqia	ENSIAS, Mohammed V University, Rabat, Morocco
Amine Saïda	FST—Mohammedia, Morocco
Amjad J. Humaidi	University of Technology, Baghdad, Iraq
Andrea Molinari	University of Trento, Italy
Arezki Fekik	Mouloud Mammeri University, Tizi Ouzou, Algeria
Aziz El Fazziki	FS Semlalia, Cadi Ayyad University, Marrakesh, Morocco
Azza Ahmed Abdo Ali	Abdulrahman Bin Faisal University, Saudi Arabia
Basma Boukenze	IR2M, FST, Hassan 1st University, Settat, Morocco
Boujemâa Achchab	ENSA, Berchid, Morocco
Brahim Ouhbi	ENSAM, Moulay Ismail University, Meknès, Morocco
Chakib Ben Njima	ENIM, Sousse University, Tunisia
Daoui Cherk	FST, Moulay Slimane University, Beni Mellal, Morocco
Driss Bouzidi	ENSIAS, Mohammed V University, Rabat, Morocco

Driss El Ouadghiri	Faculty of Sciences, Moulay Ismail University, Meknès, Morocco
El Hassan Essoufi	FST, Hassan 1st University, Settat, Morocco
El Mamoun Souidi	FS, Mohammed V University, Rabat, Morocco
El Mehdi Kandoussi	INPT, Rabat, Morocco
El Moukhtar Zemmouri	ENSAM, My Ismail University, Méknes, Morocco
Essaid Sabir	ENSEM, Hassan II University, Casablanca, Morocco
Ghizlane Orhanou	Faculty of Sciences, Mohammed V University, Rabat, Morocco
Hafssa Benaboud	Faculty of Sciences, Mohammed V University, Rabat, Morocco
Hajar Mousannif	FS Semlalia, Cadi Ayyad University, Marrakesh, Morocco
Hamid Taramit	FST, Hassan 1st University, Settat, Morocco
Hanan El Bakkali	ENSIAS, Mohammed V University, Rabat, Morocco
Hatem Ben Sta	University Tunis El Manar, Tunisia
Hicjham Ben Alla	FST, Hassan 1st University, Settat, Morocco
Hicham Tribak	Abdelmalek Essaadi University, Tétouan, Morocco
Houssam Halmaoui	ISMAL, Rabat, Morocco
Ibraheem Kasim Ibraheem	Dijlah University College, Baghdad, Iraq
Idriss Chana	EST, My Ismail University, Méknes, Morocco
Iman Elmir	I2S, Hassan 1st University, Settat, Morocco
Imane Hilal	ESI, Mohammed V University, Rabat, Morocco
Imane Lmati	FST, Hassan 1st University, Settat, Morocco
Jaouad Dabounou	FST, Hassan 1st University, Settat, Morocco
KC Santosh	University of South Dakota, USA
Karmela Aleksis Maslac	Zagreb School of Economics and Management, Croatia
Khalid El Makkaoui	Polydisciplinary Faculty, Nador, Morocco
Khalid Zine-Dine	FS, Mohammed V University, Rabat, Morocco
Krishnaveni Marimuthu	Avinashilingam University for Women, Coimbatore, India
Lavika Goel	Malaviya, National Institute of Technology (NIT), Jaipur, India
Ladjel Bellatreche	ISAE-ENSMA, Chasseneuil, France
Mehrez Abdellaoui	High Institute of Applied Sciences and Technologies, Kairouan, Tunisia
Mohamed Bahaj	FST, Hassan 1st University, Settat, Morocco
Mohamed Bakhouya	International University of Rabat, Morocco

Mohamed Chaouki Abounaima	FST, My Abdellah University, Fès, Morocco
Mohamed El Kamili	EST, Hassan II University, Casablanca, Morocco
Mohamed Et-Tolba	INPT, Rabat, Morocco
Mohamed Hanini	FST, Hassan 1st University, Settat, Morocco
Mohamed Moughit	ENSA—Khouribga, Moulay Slimane University, Morocco
Mohamed Nabil	FS, Chouaïb Doukali University, El Jadida, Morocco
Mohamed Naimi	ENSA, Berchid, Morocco
Mohamed Nemiche	Polydisciplinary Faculty—Taza, Morocco
Mohamed Sabbane	Faculty of Sciences, My Ismail University, Meknès, Morocco
Mohammed Ridounai	EST, Hassan II University, Casablanca, Morocco
Mostafa Belkasmi	ENSIAS, Mohammed V University, Rabat, Morocco
Mostafa Saadi	ENSA—Khouribga, Moulay Slimane University, Morocco
Mostafa Ezziyyani	FST, Abdelmalek Essaadi University, Tangier, Morocco
Mostapha Zbakh	ENSIAS, Mohammed V University, Rabat, Morocco
Moulay Lahcen Hasnaoui	EST, Moulay Ismail University, Meknès, Morocco
Nabil Benamar	Moulay Ismail University, Meknès, Morocco
Najib Naja	INPT, Rabat, Morocco
Najima Daoudi	ESI, Mohammed V University, Rabat, Morocco
Nabil Laachfoubi	FST, Hassan 1st University, Settat, Morocco
Najlae Idrissi	FST, Moulay Slimane University, Beni Mellal, Morocco
Nashwa Ahmad Kamal	Cairo University, Egypt
Nassereddine Bouchaib	FST, Hassan 1st University, Settat, Morocco
Nickolas S. Sapidis	University of Western Macedonia, Greece
Niketa Gandhi	MIRLabs, Washington, USA
Nizar Rokbani	University of Sousse, Tunisia
Noredine Gherabi	ENSA—Khouribga, Morocco
Ouail Ouchetto	FSJES, Université Hassan II, Casablanca, Morocco
Omar El Beqqali	Faculty of Sciences, Fès, Morocco
Parthasarathy Subashini	Avinashilingam University for Women, Coimbatore, India
Peter J. Tonelato	School of Medicine, University of Missouri, Columbia MO, USA
Rachid Dakir	Polydisciplinary Faculty, Ouarzazate, Morocco
Rachid Latif	ENSA, Ibn Zohr University, Agadir, Morocco

Rachida Ajhoun	ENSIAS, Mohammed V University, Rabat, Morocco
Rajendran Sobha Ajin Singh	Painary P.O mIdukki, Kerala, India
S. P. Raja	R&D Institute of Science and Technology, Tamil Nadu, India
Said Ben Alla	ENSA, Berchid, Morocco
Said Jai Andaloussi	FS Aïn Chock, Hassan II University, Casablanca, Morocco
Said El Kafhali	FST, Hassan 1st University, Settat, Morocco
Said Raghay	FST, Cadi Ayyad University, Marrakech, Morocco
Said Rakrak	FST, Cadi Ayyad University, Marrakech, Morocco
Salah El Hadaj	ENCG, Cadi Ayyad University, Marrakesh, Morocco
Sara Arezki	FST, Hassan 1st University, Settat, Morocco
Sofia Douda	FST, Hassan 1st University, Settat, Morocco
Umang Singh	Institute of Technology and Science, Ghaziabad, India
Youssef Balouki	FST, Hassan 1st University, Settat, Morocco
Youssef Saadi	FST, Sultan Moulay Slimane University, Beni Mellal, Morocco
Zahi Jarir	FS Semlalia, Cadi Ayyad University, Marrakesh, Morocco

Contents

Deep Learning and Intelligent Applications

Convolutional Sparse Autoencoder for Emotion Recognition	3
<i>M. Mohana and P. Subashini</i>	
Lung Cancer Classification Model Using Convolution Neural Network	16
<i>Esraa A.-R. Hamed, Mohammed A.-M. Salem, Nagwa L. Badr, and Mohamed F. Tolba</i>	
An Enhanced Deep Learning Approach for Breast Cancer Detection in Histopathology Images	27
<i>Mahmoud Ouf, Yasser Abdul-Hamid, and Ammar Mohammed</i>	
Reducing Deep Learning Complexity Toward a Fast and Efficient Classification of Traffic Signs	37
<i>Btissam Bousarhane and Driss Bouzidi</i>	
Deep Learning Approach for a Dynamic Swipe Gestures Based Continuous Authentication	48
<i>Zakaria Naji and Driss Bouzidi</i>	
Skin Cancer Detection Based on Deep Learning Methods	58
<i>Sara Shaaban, Hanan Atya, Heba Mohammed, Ahmed Sameh, Kareem Raafat, and Ahmed Magdy</i>	
Predicted Phase Using Deep Neural Networks to Enhance Esophageal Speech	68
<i>Madiha Amarjouf, Fadoua Bahja, Joseph Di-Martino, Mouhcine Chami, and El Hassan Ibn-Elhaj</i>	
State of the Art Literature on Anti-money Laundering Using Machine Learning and Deep Learning Techniques	77
<i>Bekach Youssef, Frikh Bouchra, and Ouhbi Brahim</i>	
The Reality of Artificial Intelligence Skills Among Eighth-Grade Students in Public Schools	91
<i>Shatha Sakher, Areeg Al Fouri, Shatha Al Fouri, and Muhammad Turki Alshurideh</i>	

A Deep Neural Network Architecture for Extracting Contextual Information	107
<i>Zakariae Alami Merrouni, Bouchra Frikh, and Brahim Ouhbi</i>	
Machine Learning and Applications	
Feedforward Neural Network in Cancer Treatment Response Prediction	119
<i>Hanan Ahmed, Howida A. Shedeed, Safwat Hamad, and Ashraf S. Hussein</i>	
A Genetic Algorithm Approach Applied to the Cover Set Scheduling Problem for Maximizing Wireless Sensor Networks Lifetime	129
<i>Ibtissam Larhlimi, Maryem Lachgar, Hicham Ouchitachen, Anouar Darif, and Hicham Mouncif</i>	
Application of Machine Learning to Sentiment Analysis	138
<i>Oumaima Bellar, Amine Baina, and Mostafa Bellafkih</i>	
Robust Vehicle Detection by Using Deep Learning Feature and Support Vector Machine	149
<i>Vinh Dinh Nguyen, Thanh Hoang Tran, Doan Thai Dang, and Narayan C. Debnath</i>	
Arabic Vowels Recognition Using Envelope's Energy and Artificial Neural Network	158
<i>Nesrine Abajaddi, Youssef Elfahm, Badia Mounir, and Abdelmajid Farchi</i>	
Soil Nutrient Prediction Model in Hybrid Farming Using Rule-Based Regressor	164
<i>M. Krishnaveni, Paa. Raajeswari, P. Subashini, V. Narmadha, and P. Ramya</i>	
Evaluation and Comparison of Energy Consumption Prediction Models Case Study: Smart Home	179
<i>Elhabyb Khaoula, Baina Amine, and Bellafkih Mostafa</i>	
Leaf Disease Detection in Blueberry Using Efficient Semi-supervised Learning Approach	188
<i>Vinh Dinh Nguyen, Ngoc Phuong Ngo, and Narayan C. Debnath</i>	
Student Performance Prediction in Learning Management System Using Small Dataset	197
<i>Zakaria Soufiane Hafdi and Said El Kafhali</i>	

Modeling of Psuedomorphic High Electron Mobility Transistor Using Artificial Neural Network	206
<i>Radwa Mohamed, Ahmed Magdy, and Sherif F. Nafea</i>	

Leveraging Blockchain and Machine Learning to Improve IoT Security for Smart Cities	216
<i>Mayar M. Moawad, Magda M. Madbouly, and Shawkat K. Guirguis</i>	

Image Processing and Computer Vision

Red-Channel Based Iris Segmentation for Pupil Detection	231
<i>S. Bhuvaneswari and P. Subashini</i>	

Low Overhead Routing in a Lightweight Routing Protocol	242
<i>Maryem Lachgar, Ibtissam Larhlimi, Anouar Darif, Hicham Ouchitachen, and Hicham Mouncif</i>	

Identification and Localization COVID-19 Abnormalities on Chest Radiographs	251
<i>Van Tien Pham and Thanh Phuong Nguyen</i>	

Breathing Pattern Assessment Through the Empirical Mode Decomposition and the Empirical Wavelet Transform Algorithms	262
<i>Zakaria El Khadiri, Rachid Latif, and Amine Saddik</i>	

Lung Cancer Stages Classification Based on Differential Gene Expression	272
<i>Moshira S. Ghaleb, Hala M. Ebied, and Mohamed F. Tolba</i>	

Deep Learning for Image and Sound Data: An Overview	282
<i>Hilali Manal, Ezzati Abdellah, and Ben Alla Said</i>	

Correction of an Image for Colour Blindness Using the Fusion of Ishihara Filter and Histogram Equalization	294
<i>M. S. Sannidhan, Jason Elroy Martis, C. V. Aravinda, and Roheet Bhatnagar</i>	

Color Image Cryptography Using Block and Pixel-Wise Permutations with 3D Chaotic Diffusion in Metaverse	305
<i>Renjith V. Ravi, Pushan Kumar Dutta, and Sudipta Roy</i>	

Sentiment Analysis and Recommendation Systems

Review on Recent Trends in Recommender Systems for Smart Cities	317
<i>Sana Abakarim, Sara Qassimi, and Said Rakrak</i>	

Sentiment Analysis for Competence-Based e-Assessment Using Machine Learning and Lexicon Approach	327
<i>Mohammed Amraouy, Mostafa Bellafkih, Abdellah Bennane, and Jallal Talaghzi</i>	
User Sentiment Analysis Towards Adapting Smart Cities in Egypt	337
<i>Lamiaa Mostafa and Sara Beshir</i>	
Toward a Generative Chatbot for an OER Recommender System Designed for the Teaching Community: General Architecture and Technical Components	348
<i>Sandoussi Rima, Hnida Meriem, Daoudi Najima, and Ajhoun Rachida</i>	
An Emoticon-Based Sentiment Aggregation on Metaverse Related Tweets	358
<i>Mousumi Bhattacharyya, Asmita Roy, Sadip Midya, Anirban Mitra, Anupam Ghosh, and Sudipta Roy</i>	
A Personalized Recommender System Based-on Knowledge Graph Embeddings	368
<i>Ngoc Luyen Le, Marie-Hélène Abel, and Philippe Gouspillou</i>	
Awareness of Arabic Teachers in Academic Integrity Towards Students Completing Tasks in Abu Dhabi: Data Analysis Approach	379
<i>Suad Abdalkareem Alwaely, Asaad Ali Muslim Al-Maamari, and Muhammad Turki Alshurideh</i>	
Data Sciences and Business Based Applications	
Short-Term Forecasting of GDP Growth for the Petroleum Exporting Countries Based on ARIMA Model	399
<i>Sara Abdelghafar, Ashraf Darwish, and Abdulrahman Ali</i>	
Technique for Order of Preference by Similarity to Ideal Solution Based Methodology for Detecting Important Actors in Social Networks: Facebook Ego Network as a Case Study	407
<i>Khaoula Ait Rai, Mustapha Machkour, and Jilali Antari</i>	
Evaluating FFT-Based Convolutions on Skin Diseases Dataset	416
<i>Amina Aboulmira, Hamid Hrimch, and Mohamed Lachgar</i>	
On the Detection of Anomalous or Out-of-Distribution Data in Vision Models Using Statistical Techniques	426
<i>Laura O'Mahony, David JP O'Sullivan, and Nikola S. Nikolov</i>	

Autonomous Vehicle Algorithm Decision-Making Considering Other Road Users	436
<i>Ishan Tyagi and Subramaniam Ganesan</i>	
The Role of Administrative Leadership in Crisis Management in the Jordanian Ministry of Planning and International Cooperation	447
<i>Haron Ismail Al-lawama, Saddam Rateb Darawsheh, Mohammad Salameh Zaid Almahairah, Anwar Saud Al-Shaar, Asmaa Jumaha AlMahdawi, Azhar Shater, Imen Gmach, Hamed Omar Abdalla, and Muhammad Turki Alshurideh</i>	
Remote Working in the COVID-19 Era	459
<i>Randa Abu Hamour, Areeg Alfouri, and Muhammad Alshurideh</i>	
Metaheuristic Algorithms-Based Applications	
Metaheuristic Algorithms Based Server Consolidation for Tasks Scheduling in Cloud Computing Environment	477
<i>Hind Mikram, Said El Kafhali, and Youssef Saadi</i>	
Spectrum Recovery Improvement in Cognitive Radio Using Grey Wolf Optimizer	487
<i>Gehan Gamal, Mohamed F. Abdelkader, Abdelazeem A. Abdelsalam, and Ahmed Magdy</i>	
An Efficient Meta-Heuristic Methods for Travelling Salesman Problem	498
<i>Mohamed Abid, Said El Kafhali, Abdellah Amzil, and Mohamed Hanini</i>	
Optimization of Task Scheduling in Cloud Computing Using the RAO-3 Algorithm	508
<i>Ahmed Rabie Fayed, Nour Eldeen M. Khalifa, M. H. N. Taha, and Amira Kotb</i>	
Multilevel Quantum Evolutionary Butterfly Optimization Algorithm for Automatic Clustering of Hyperspectral Images	524
<i>Tulika Dutta, Siddhartha Bhattacharyya, and Bijaya Ketan Panigrahi</i>	
Enhancement of Security in GFDM Using Ebola-Optimized Joint Secure Compressive Sensing Encryption and Symbol Scrambling Model	535
<i>Irfan Ahmad Rather, Gulshan Kumar, Rahul Saha, and Tai-hoon Kim</i>	

Software-Defined Network and Telecommunication

CPC-SDN: A Centralized Proactive Caching Based on Software Defined Network 553
Sembati Yassine, Naja Najib, and Jamali Abdellah

New Approach to Telecom Churn Prediction Based on Transformers 565
Jalal Rabbah, Mohammed Ridouani, and Larbi Hassouni

Smart Healthcare Development Based on IoMT and Edge-Cloud Computing: A Systematic Survey 575
Fatima Ezzahra Moujahid, Siham Aouad, and Mostapha Zbakh

The Impact of Leadership on Employee Motivation in the Jordanian Telecommunication Sector 594
Haron Ismail Al-lawam, Mohammad Salameh Zaid Almahairah, Hiba Hussein Mohammad Almomani, Ahmad A. I. Shajrawi, Saddam Rateb Darawsheh, Anwar Saud Al-Shaar, and Muhammad Turki Alshurideh

Author Index 605

Deep Learning and Intelligent Applications



Convolutional Sparse Autoencoder for Emotion Recognition

M. Mohana^(✉) and P. Subashini

Department of Computer Science, Centre for Machine Learning and Intelligence (CMLI),
Avinashilingam Institute, Coimbatore, India
{mohana_cs, subashini_cs}@avinutty.ac.in

Abstract. Emotion recognition is a hot research area in deep learning and computer vision that analyses expressions from both static and dynamic sequences of facial expressions to reveal human emotional states. In recent decades, deep learning approaches have been exhibiting a superior performance on image representation datasets. However, the convolutional neural network (CNN) requires a larger number of labeled datasets for training and accurate classification results. It is always inevitable, whereas unsupervised representation learning models like autoencoder do not require labeled information for training. Meanwhile, it is difficult to infer the feature map when the size of the CNN layer is increased. To address these challenges, this paper introduced a self-supervised deep learning technique called convolutional sparse autoencoder (CSA) which can learn robust features from small data with unlabeled facial expression datasets. Moreover, sparsity is added in the max pooling layer for the feature map which makes the backpropagation optimizer Adam work efficiently for the CSA training; thus, no complicated optimizer is not involved. Finally, the trained convolutional sparse encoder part is combined with the softmax layer for emotion classification. The performance results demonstrate that the proposed approach achieved 98% of accuracy on the CK+ dataset and outperforms various state-of-the-art methods.

Keywords: Convolutional sparse autoencoder (CSA) · Convolutional neural network (CNN) · Deep learning · Emotion recognition (ER) · Representation learning (RL)

1 Introduction

Emotion recognition has gained popularity over the past decades of its applications [1] in a variety of fields, including the development of intelligent robots, improved emotional understanding, monitoring candidates' expressions during interviews, and so on. Furthermore, it is an essential component of human-robot interaction technology. Humans find it easy to recognize human emotions, but it is difficult for machines. Numerous research has been conducted in the literature, and a few emotions are mentioned as universal expressions that are happy, sad, fear, angry, surprise, and disgust while some studies added neutral and disgust. Furthermore, many challenges such as illumination,

pose variation, low resolution, and complex background always have issues with facial expression recognition performance.

There are different styles of approaches and techniques applied to facial emotion recognition like other recognition research. The primary function of FER is to map an expression to the appropriate emotional state. Face detection, pre-processing, feature extraction, and emotion recognition are the four steps in the conventional FER. Except for facial images, the face detection method is used to identify the face in images and remove the background parts. The next pre-processing steps include resizing, cropping, and normalization. Existing works use notable approaches such as histogram-oriented gradients (HOG), scale-invariant feature transform (SIFT), facial action units, and Gabor wavelet transform to extract features for the conventional FER system. Moreover, dimension reduction techniques, such as local binary pattern (LBP), principal component analysis (PCA), and optical flow techniques, also help to extract relevant descriptors from facial images. Prior to the introduction of deep learning techniques in the FER system, those approaches aided in removing irrelevant information from facial images and improving recognizing ability. Conventional feature extraction approaches, however, are insufficient for extracting micro-expressions from facial images.

Representation learning is an effective technique for learning high-level feature representations from images [4, 6]. Convolutional neural networks (CNN) in deep learning are typically applied in numerous image recognition and detection processes [7]. The CNN structure and pooling are important factors in representation learning, and the convolutional layer is used to learn spatial information from static images. The pooling layer is used to decrease the dimension of the features and overfitting issues [13]. However, CNN requires a large number of labeled data when training the model, resulting in high computational costs. Furthermore, the layer size is increasing while the network is becoming deeper for extracting high-level features, which frequently results in a vanishing gradient problem.

To address these challenges, this paper introduced the convolutional sparse autoencoder (CSA) for emotion classification. The presented method is first trained with an unsupervised small amount of facial expression dataset, as illustrated in Fig. 3. Following that, the trained encoder has been combined with the softmax layer and fine-tuned for emotion classification. Here, sparsity [11] has been added to each max pooling layer to enhance and speed up the feature map for efficient feature learning [12]. The outcome of the experiments demonstrates that the presented method is effective for image feature extraction using a data-driven approach and achieves robust labeling for facial expression recognition. This paper's main contribution is as follows:

1. CSA is proposed for the automatic extraction of features from small amounts of data while avoiding the uncertainty of the conventional feature selection process. This technique is effective for dealing with the problem of insufficient data training caused by labeled facial expression images,
2. With the help of the robust power of convolutional kernels, facial traits are directly extracted from raw images. Moreover, this method reduced the size of parameters and overfitting issues compared to the conventional approach, and
3. Any type of image dataset and associated application domains can be extracted using the unsupervised data-driven approach.

The structure of the paper is as follows: Sect. 2 describes the already existing works based on convolutional neural networks and autoencoder on the FER system. Section 3 explains the proposed approach step by step. Section 4 discusses the results and analyses of the proposed network on CK+ datasets. Finally, Sect. 5 summarizes the overall CSA work.

2 Background and Related Works

During recent decades, deep learning techniques, particularly in FER, have excelled at automatic image recognition and detection tasks [4, 17]. The conventional method extracts feature from facial images before classifying emotions based on feature values. Through the evaluation of the FER system, various machine learning methods (K-nearest neighbor, Support vector machine, neural network, principal component analysis, local binary pattern, and linear discriminant analysis) are initially employed. Such techniques have the limitation of extracting features from front views of facial images for FER evaluation. Deep learning techniques based on FER tasks, on the other hand, combine feature selection and emotion classification processes. Numerous studies have reviewed and compared [4–7], and the recent unsupervised learning-based FER is also included [14, 16]. The techniques used in eminent FER methods are briefly described in the studies that follow.

Zhao et al. [4] have proposed the convolutional neural network-based FER system and fine-tuned it with the VGG network. This network has been trained with different sizes of convolutional filters and dropout values to generalize the network. The authors have achieved the performance of this network at 99.33% on CK+, 87.65% on MUG, and 93.33% on RaFD datasets. Mollahosseini et al. [5] introduced deep learning-based architecture to address the issues in FER on different datasets. Two convolutional networks and a pooling layer were used to create this network, which was then followed by inception layers that increased the depth and width of the network during keeping the computing cost stable. It is a single-network architecture that accepts facial images as input and categorizes them based on one of six primary emotions. These experiments were conducted on CK+, DISFA, FERA, SFEW, MultiPIE, MMI, and FER 2013. Minaee et al. [7] proposed a deep network for improving the accuracy of multiple datasets, including FERG, JAFFE, FER-2013, and CK+, based on an attentional convolutional neural network that focused on important features of facial images. This network has achieved some extent performance compared to conventional FER techniques. Jaiswal et al. [6] demonstrated emotion recognition using a convolutional neural network that combines two network features to classify emotions. It has been tested on two different datasets with 70.14% accuracy on FER-2013, and 98.65% on JAFFE. Akhand et al. [8] have introduced a deep convolutional network (DCN) through transfer learning (TL) techniques and fine-tuned it with facial expression data. The limitation of the conventional CNN network on the facial image is features extracted from the frontal view of high-resolution images. For this experiment, the authors include eight different pre-trained networks (ResNet-18, ResNet-34, ResNet-50, ResNet-152, Inception-v3, VGG-16, VGG-19, and DenseNet-161) for improving the accuracy of the FER system. This method experimented on JAFFE and KDFE datasets and achieved 96.51%, and 99.52% respectively.

Unsupervised artificial neural network-based autoencoders is widely used in several real-world applications like anomaly detection, data compression, image denoising,

segmentation, and classification in recent decades [16]. Zeng et al. [14] used a deep sparse autoencoder to learn robust features from unlabeled facial image data to recognize facial expressions with high accuracy. Facial features were extracted using both geometric and appearance features instead of extracting features from raw facial images. This technique helped to identify the relevant features. These experimental results show 95.79% accuracy was achieved on CK+ datasets. Liu et al. [15] proposed a stacked sparse autoencoder that uses an optical flow method combined with a deep neural network to reduce the influence of the same expression on different facial expressions. Finally, a softmax layer had been added on top of the layer for emotion classification. This work has experimented on CK+ datasets and achieved 92.3% accuracy. Usman et al. [16] have introduced emotion recognition using deep sparse autoencoder for feature selection and dimensional reduction for facial expression recognition on multiple hidden layers. The authors showed that the stacked autoencoder extracted more relevant features from facial images compared to the conventional approach and achieved 99.60% accuracy on CK+ datasets. The dense network is used in the works mentioned above to recognize facial emotions. Zhang [21] has compared the performance of simple autoencoder and convolutional autoencoder for image pre-processing. The results show that the convolutional autoencoder gives better performance on the image dataset.

3 Convolutional Sparse Autoencoder for Emotion Recognition

The proposed emotion recognition framework uses a convolutional sparse autoencoder (CSA) to extract the latent representation for each class from facial images and a softmax layer to perform classification. This methodology is divided into the following steps: pre-processing, convolutional sparse autoencoder, and emotions classification. Figure 1 shows the proposed network architecture.

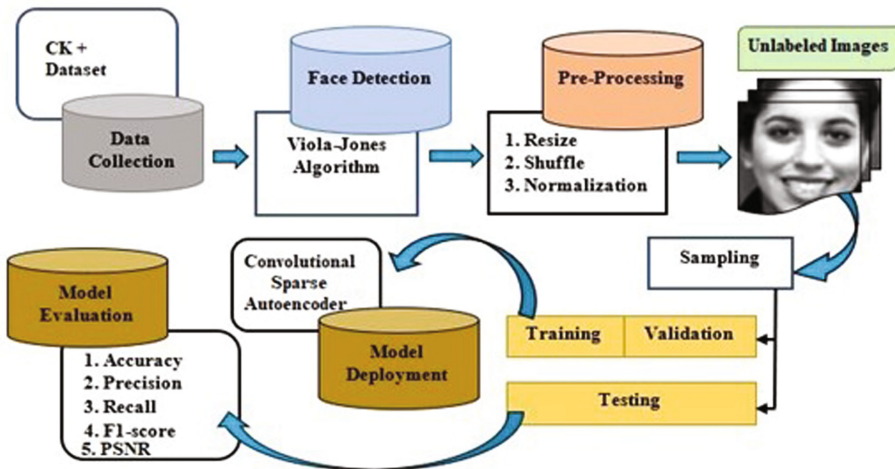


Fig. 1. Overview of the proposed architecture

3.1 Pre-processing

Pre-processing is a critical stage in image processing that is used to improve the relevant features for subsequent steps. The first process of a conventional FER system is face detection which is used to identify and locate the face in images. For facial emotion recognition, the facial part is more sufficient than other parts of features in facial images. For this purpose, a real-time viola-jones face [9] detection algorithm has been employed in this research study to detect the face from raw images (See Fig. 2). This algorithm consists of four parts, namely, haar features, integral images, Adaboost algorithm, and cascade classifiers. Firstly, Haar features consist of black and white regions which produce a single value by the sum of the light regions subtracted by the sum of the black regions. It is used to extract useful information such as edges, and diagonal and straight lines for identifying the human face. Next integral images are used to speed up the haar rectangle feature calculation process. Third, the AdaBoost algorithm is used to build a strong classifier among all available features. Finally, the cascade classifier removed the unnecessary part from facial images except for the facial region. In addition, the detected face has cropped into 48 x 48 pixels whose intensity values have been normalized between 0–1 for reducing the computation time complexity in a neural network.



Fig. 2. Face detection process

3.2 Convolutional Sparse Autoencoder

A supervised method is a data-driven feature learning method that updates connection weights through forward and backward training processes. Unsupervised learning directly receives unlabeled input data and learns more relevant features compared to the supervised method. This method significantly reduced the workload for labeling data. Figure 3 shows the overall structure of the proposed method.

In this paper, the unsupervised autoencoder [3], which is made up of three parts: encoder, latent space (code), and decoder, effectively recognizes facial expressions. The autoencoder is primarily used for dimensionality reduction, image denoising, and feature extraction. It gives an output that is the same as the input data and compares its original data. After many iterations, the cost function reaches optimality, which means that the reconstructed output is as close to the input data as possible. The encoder converts the input data into code of the hidden layer by $code(c) = f(w.x + b)$, where f is an activation function, w is a weight x is an input value and b is a bias. The decoder

reconstructs the output from code of the hidden layer by $x' = f'(w'.c + b')$, and calculate the mean squared values between input and reconstructed output using the cost function $cost = \min \sum_{i=1}^n |x' - x|^2$.

A convolutional autoencoder [2, 18] is a variational of a convolutional neural network that is used for retaining the connected information between pixels of images. The layers in the CNN network help to extract the relevant features and find patterns without human intervention. The process of converting the feature maps input to output is known as a convolutional encoder, and the output is reconstructed using the inverse convolutional operation, which is known as a convolutional decoder. Moreover, the reconstruction error of the convolutional encoder and decoder can be calculated in the same way as the standard autoencoder.

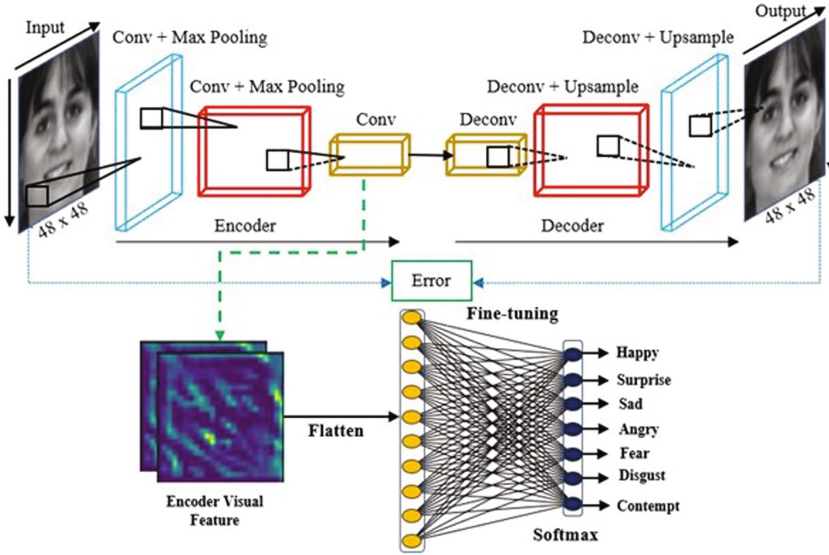


Fig. 3. Proposed convolutional sparse autoencoder (CSA)

The number of convolutional layers, pooling layer, ReLU activation function, neuron size, sparsity, and filter size are the main components of a convolutional sparse autoencoder. The proposed model consists of four convolution blocks of the encoder and three deconvolution blocks of the decoder, each with a convolutional layer and filter size (3×3) , followed by batch normalization, which is used to generalize the network learning process. After each block, the max pooling (a kernel size 2×2) layer with sparsity [12] is used to downsample the feature map of the convolutional encoder. Here, the sparsity helps to highlight the more relevant features for learning. In the convolutional decoder, the convolutional layer, activation function, and batch normalization are also included. After the two convolutional blocks, upsampling layers (kernel size 2×2) are used.

First, the input facial image is encoded each time with pixels patch $x_i, i = 1, 2 \dots x_n$ and multiplied with neuron weights w_j , where j is used for convolutional calculation. Finally, the output layer o_{ij} is calculated as $o_{ij} = f(w_j.x_i + b)$. Then, output from the convolutional decoder is defined as $x'_i = f'(w'_j.o_{ij} + b')$. Finally, reconstructed

error is calculated as $CSA = \frac{1}{P} \sum_{i=1}^P \|x_i - x'_i\|$, where p is reconstruction operation of convolutional kernel size with $d \times d$, where $d \leq pixels$.

3.3 Emotion Classification

In the classification part, the last convolutional layer in the encoder part has flattened into a feature vector and fed into a dense layer. A flattened layer is followed by a dense layer containing 128 neurons and a ReLU activation function. For more than one emotion classification, the softmax layer has added the top of the convolutional encoder. In the training process, the encoder has been fine-tuned by a softmax layer with a labeled facial expression dataset after the training of convolution sparse autoencoder. It has been trained with 100 epoch 32 batch size and Adam optimizer with 0.001 learning rate. In addition, categorical cross entropy is used as a loss function for calculating the training and validation accuracy of the presented model performance.

4 Experimental Results and Discussion

4.1 Datasets

For this research study, the CK+ [25] database is used for CSA performance evaluation. This dataset, which was released in 2010, is an expanded version of the Cohn-Kanade (CK), one of the most widely used benchmark datasets for evaluating the FER algorithm. The CK+ dataset contains 593 video sequences from 123 subjects at 30 frames per second and 640×490 pixels in resolution, which includes eight basic facial expressions such as 276 happy, 180 anger, 112 sad, 332 surprises, 72 contempt, 236 disgust, 100 fear and 327 neutral used for this evaluation. The length of each video sequence ranges from 10 to 60 frames.

In addition, the proposed approach has been designed using Keras and TensorFlow backend. The hold-out cross-validation method is used in this study, with 80% of the facial images used for training and the remaining 20% used for testing.

4.2 Evaluation Metrics

The following metrics are used in this experiment to assess the performance of the presented network emotion classification. FN stands for False Negative, TP stands for True Positive, FP stands for False Positive, and TN stands for True. To summarize the performance of the classification algorithm is called a confusion matrix from which values of performance are calculated individually.

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN} \quad (1)$$

Accuracy is the proportion of correctly classified instances to the total number of instances. It can judge positive as positive and negative as negative.

$$Recall/Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

Recall refers to the predicted positive cases from the total positive cases.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Precision refers to the number of positive predictions in the positive example classified by the detection system.

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \quad (4)$$

F1-Score is a harmonic mean of combining the values of precision and recall.

$$PSNR = 10 \log_{10} \frac{Max_I^2}{MSE} \quad (5)$$

The MSE is used to calculate the autoencoder's reconstruction loss, whereas the peak signal-to-noise ratio (PSNR) is used to compare the quality of the original and compressed reconstructed images.

4.3 Experimental Results

This section presents the proposed network performance on test data and compares it with existing works. This network consists of four convolutional blocks of the encoder and three deconvolutional blocks of the decoder. The sparsity parameter is set to $1e-5$. Initially, convolutional sparse autoencoder trained with 100 epochs, 32 batch size and Adam optimizer with learning rate 0.001. Figure 4. Shows the loss and reconstruction performances of CSA. The training loss for the proposed network is 0.0228, and the validation loss is 0.0428. Furthermore, MSE and PSNR are two more commonly used metrics. MSE's limitation is that it is highly dependent on image intensity scaling. PSNR prevents these issues by scaling the MSE according to the image range, and the reconstructed facial expression images achieved 70.06 dB PSNR in the presented network.

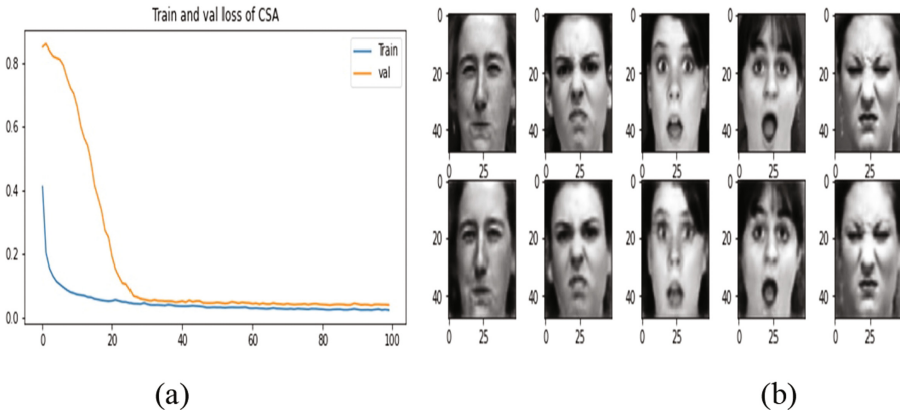


Fig. 4. (a) Loss evaluation of CSA (b) Sample test image, the first row shows the original images, and the second row shows the reconstructed images

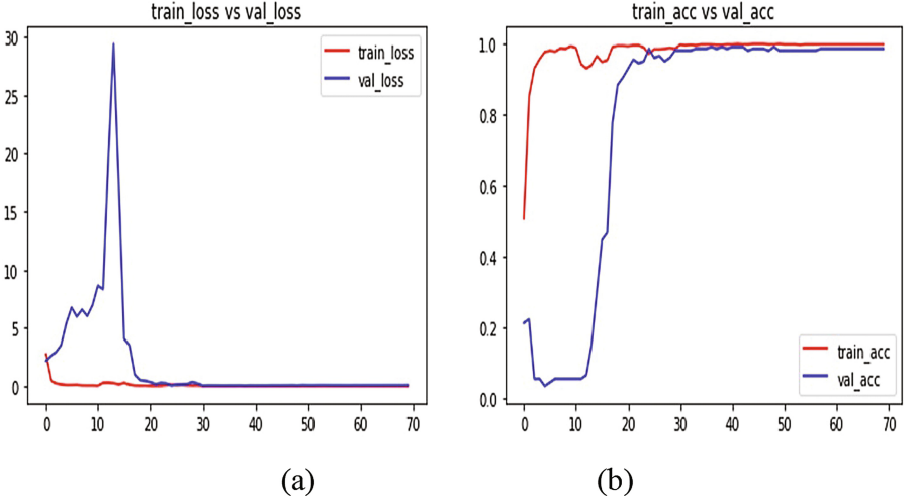


Fig. 5. (a) Training and validation loss of CSA (b) Training validation accuracy of CSA

After training the entire network, the final convolutional encoder layer is flattened into a feature vector and fed into the dense layer. Then, on top of the convolution encoder, a softmax layer with four convolution blocks is added. This network is trained using 70 epochs, 32 batch sizes, and the Adam optimizer with a learning rate of 0.001. Figure 5. Depicts CSA's training and testing performance. The training loss and accuracy, and validation loss and accuracy have been recorded in every epoch. Initially, the loss value is gradually increased due to the random initialization of weights. After the 30th epoch, validation loss decreased step by step and reached a minimum loss of about 0.0726. In the process of backpropagation, the encoding weights of each layer can be improved further for the ability of feature extraction. In Fig. 5. (b), initially, the network starts with higher fluctuations due to imbalanced facial expression images, after the optimizer Adam helped to generalize the network training. Finally, on the CK+ dataset, the presented model achieved 0.98% accuracy, 0.96% precision, 0.98% recall, and 0.97% f1-score. Furthermore, the accuracy of the seven classes is depicted in Fig. 6. The distinct features surrounding the eyes, lips, and nose of the region helped contempt, disgust, happy, and sad reach a greater performance accuracy of 100% among the seven expressions from this depiction. Here anger, fear, and surprise are slightly misclassified as contempt with 0.04%, 0.1%, 0.02% respectively. In general, facial expressions are readily mistaken due to their similar shape and look as well as their distinct variations within the same expression. In addition, table demonstrates how the suggested approach outperforms the other five FER algorithms, as can be observed. The reason for choosing these methods tested on the same CK+ datasets with outstating different approach and performances. From this comparison, convolutional sparse autoencoder discriminate different features and reduce the high dimension features, which also acts as robust facial feature classifier.

Visualizing the information captured behind the CNN network is highly important for evaluating the performance of the model. For this Grad-CAM technique is used

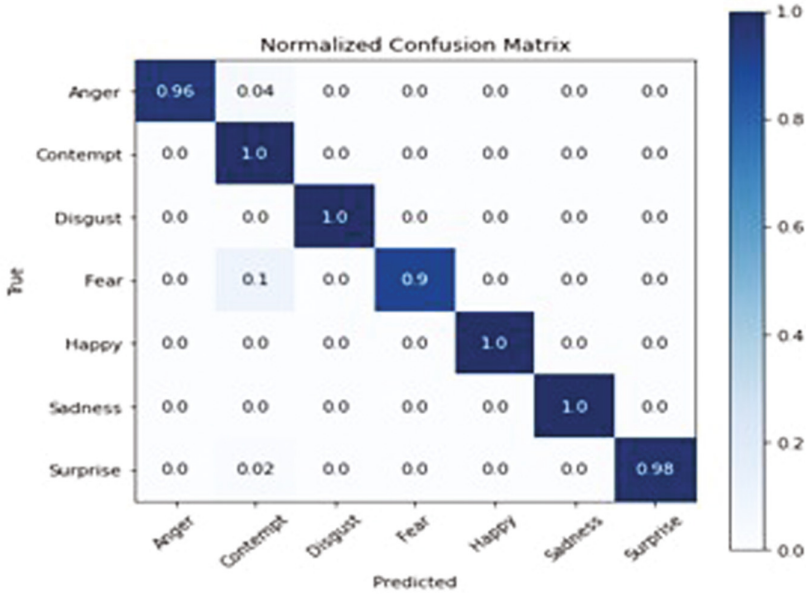


Fig. 6. Confusion matrix of 7 class facial expression recognition on CK+ dataset

here to differentiate between and capture facial emotions. Figure 7 shows the Grad-CAM visualization. The presented model focuses on important aspects of the image, i.e., lips, eyes, and eyebrows which help to distinguish the different emotions. Furthermore, Table 1 shows the proposed method recognition accuracy compared with state-art-of-the techniques results on each expression. The proposed method results are relatively high and reasonable on each facial expression.

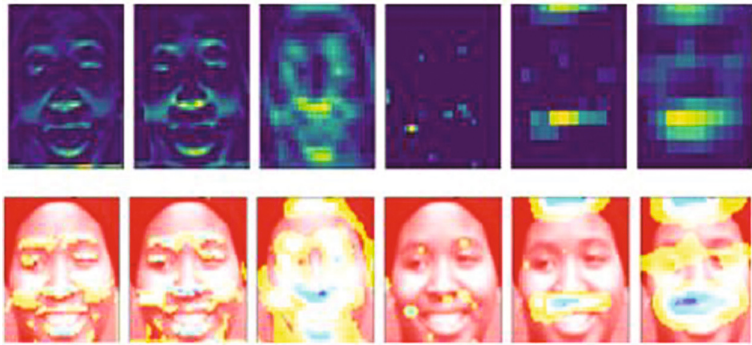


Fig. 7. Grad-CAM visualization [24] of sample emotion

Table 1. Performance evaluation of CSA versus other FER techniques

%	[19]	[20]	[21]	[22]	[23]	Proposed (CSA)
Anger	100	82.50	87.1	87.8	87.00	96.00
Contempt	83.34	–	–	–	–	100
Disgust	91.52	97.50	90.20	93.33	83.00	100
Fear	88.00	95.00	92.00	94.33	89.00	90.00
Happy	100	100	98.07	94.20	90.00	100
Sadness	85.51	92.50	91.47	96.42	84.00	100
Surprise	95.18	92.50	100	98.46	90.00	98.00
Average	91.94	93.33	93.14	94.09	87.16	98.00

5 Conclusion

A convolutional sparse autoencoder has been proposed in this paper to recognize emotions from facial expressions. This model is fully automated without the need for manual feature extraction. This autoencoder's primary goal is to reduce the issues of overfitting and the large amount of labeled data needed in the conventional FER method. Initially, it has been trained with unsupervised facial expressions. After that convolutional encoder part is trained and fine-tuned by the labeled facial expression dataset. The softmax layer is used to classify each emotion according to features learned by the encoder. Different learning rates, epochs, and batch sizes have been tried on this CSA for getting better configuration hyperparameters. Finally, the proposed model achieved 98% accuracy with a 0.0726 validation loss. Furthermore, this approach has been validated with precision, recall, F1-score, and PSNR metrics. Additionally, this experimental finding has been analyzed with existing state-of-the-art techniques. In the future, for emotion recognition, the physiological signal will be combined with facial expressions to handle the challenges of the real-world environment.

Acknowledgment. The authors wish to express their sincere thanks to the Centre for Machine Learning and Intelligence (CMLI) for providing resources to conduct this research study. This centre is sponsored and supported by the Department of Science and Technology (DST)-CURIE, India.

References

1. Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W., Wróbel, M.R.: Emotion recognition and its applications. In: Hippe, Z., Kulikowski, J., Mroczek, T., Wtorek, J. (eds.) Human-Computer Systems Interaction: Backgrounds and Applications 3. AISC, vol. 300, pp. 51–62. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08491-6_5

2. Zhang, Y.: A better autoencoder for image: convolutional autoencoder. In: ICONIP17-DCEC (2018). http://users.cecs.anu.edu.au/Tom.Gedeon/conf/ABCs2018/paper/ABCs2018_paper_58.pdf
3. Bank, D., Koenigstein, N., Giryas, R.: Autoencoders. arXiv preprint [arXiv:2003.05991](https://arxiv.org/abs/2003.05991) (2020)
4. Zhao, X., Shi, X., Zhang, S.: Facial expression recognition via deep learning. *IETE Tech. Rev.* **32**(5), 347–355 (2015)
5. Mollahosseini, A., Chan, D., Mahoor, M.H.: Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–10. IEEE (2016)
6. Jaiswal, A., Raju, A.K., Deb, S.: Facial emotion detection using deep learning. In: 2020 International Conference for Emerging Technology (INCET), pp. 1–5. IEEE (2020)
7. Minaee, S., Minaei, M., Abdolrashidi, A.: Deep emotion: facial expression recognition using the attentional convolutional network. *Sensors* **21**(9), 3046 (2021)
8. Akhand, M.A.H., Roy, S., Siddique, N., Kamal, M.A.S., Shimamura, T.: Facial emotion recognition using transfer learning in the deep CNN. *Electronics* **10**(9), 1036(2021)
9. Viola, P., Jones, M.J.: Robust real-time face detection. *Int. J. Comput. Vis.* **57**(2), 137–154 (2004)
10. Kavukcuoglu, K., Sermanet, P., Boureau, Y.L., Gregor, K., Mathieu, M., Cun, Y.: Learning convolutional feature hierarchies for visual recognition. In: *Advances in Neural Information Processing Systems*, vol. 23 (2010)
11. Bristow, H., Eriksson, A., Lucey, S.: Fast convolutional sparse coding. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 391–398 (2013)
12. Rigamonti, R., et al.: On the relevance of sparsity for image classification. *Comput. Vis. Image Underst.* **125**, 115–127 (2014)
13. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014. LNCS*, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
14. Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., Dobaie, A.M.: Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing* **273**, 643–649 (2018)
15. Liu, Y., Hou, X., Chen, J., Yang, C., Su, G., Dou, W.: Facial expression recognition and generation using sparse autoencoder. In: 2014 International Conference on Smart Computing, pp. 125–130 (2014). IEEE
16. Usman, M., Latif, S., Qadir, J.: Using deep autoencoders for facial expression recognition. In: 2017 13th International Conference on Emerging Technologies (ICET), pp. 1–6 (2017). IEEE
17. Lv, Y., Feng, Z., Xu, C.: Facial expression recognition via deep learning. In: 2014 International Conference on Smart Computing, pp. 303–308 (2014). IEEE
18. Masci, J., Meier, U., Cireşan, D., Schmidhuber, J.: Stacked convolutional auto-encoders for hierarchical feature extraction. In: Honkela, T., Duch, W., Girolami, M., Kaski, S. (eds.) *ICANN 2011. ICANN 2011. LNCS*, vol. 6791, pp. 52–59. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21735-7_7
19. Boughida, A., Kouahla, M.N., Laffi, Y.: A novel approach for facial expression recognition based on Gabor filters and genetic algorithm. *Evol. Syst.* **13**(2), 331–345 (2021)
20. Uddin, M.Z., Lee, J.J., Kim, T.S.: An enhanced independent component-based human facial expression recognition from video. *IEEE Trans. Consum. Electron.* **55**(4), 2216–2224 (2009)
21. Zhang, L., Tjondronegoro, D.: Facial expression recognition using facial movement features. *IEEE Trans. Affect. Comput.* **2**(4), 219–229 (2011)
22. Happy, S.L., Routray, A.: Automatic facial expression recognition using features of salient facial patches. *IEEE Trans. Affect. Comput.* **6**(1), 1–12 (2014)

23. Mishra, S., Joshi, B., Paudyal, R., Chaulagain, D., Shakya, S.: Deep residual learning for facial emotion recognition. In: Shakya, S., Bestak, R., Palanisamy, R., Kamel, K.A. (eds.) *Mobile Computing and Sustainable Informatics. LNDECT*, vol. 68, pp. 301–313. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-1866-6_22
24. Yang, S., Kim, Y., Kim, Y., Kim, C.: Combinational class activation maps for weakly supervised object localization. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2941–2949 (2020)
25. The Extended Cohn-Kanada Database. <https://www.ri.cmu.edu/>. Accessed 15 Nov 2022



Lung Cancer Classification Model Using Convolution Neural Network

Esraa A.-R. Hamed¹(✉), Mohammed A.-M. Salem², Nagwa L. Badr¹,
and Mohamed F. Tolba¹

¹ Faculty of Computer and Information Sciences, Ain Shams University Cairo, Cairo, Egypt
{esraa.raoof, nagwabadr, fahmytolba}@cis.asu.edu.eg

² Media Engineering and Technology, GUC, Cairo, Egypt
mohammed.salem@guc.edu.eg

Abstract. Lung cancer has one of the worst mortality rates and incidence rates of any prevalent cancers around the world, and sufferers have a better chance of surviving if the illness is detected early. One of the elements essential to determining the type of cancer is the histopathological diagnosis. Because the type of histology, molecular profile, and stage of the cancer all affect how the disease is treated, it is urgently necessary to analyze the histopathology images of lung cancer. Therefore, deep learning techniques are used to speed up the crucial process of diagnosing lung cancer and lessen the workload on pathologists. It focuses on giving computers the ability to perceive, recognize, and process images in a manner similar to that of human vision and subsequently produce the intended results. It is comparable to imparting human intelligence and instincts to a computer. Deep learning techniques, particularly Convolutional Neural Network (CNN), have improved efficiency in the analysis of cancer histopathology slides. The novelty of this work is to investigate the effectiveness of the proposed model in classifying the digital pathology images for those lung cancer images as either benign or malignant. From the LC25000 dataset, which includes 5000 images for each class, a total of 10,000 digital images were obtained. The histological slides have been divided into benign and malignant cells (cancerous cell squamous) using a shallow neural network structure. The applied model has been attained accuracy ranging from 99.3% to 99.8% in classifying lung malignant from benign lesions. It was experimentally proved that there is no tangible difference in average accuracy between the applied experiments. Therefore, the proposed model has been efficient in comparison with the state-of-the-art.

Keywords: Lung cancer · Histopathological · Squamous cell carcinomas · Convolutional neural networks · Deep learning

1 Introduction

Cancer is the leading cause of mortality for nearly 10 million deaths in 2020 in the world, according to the World Health Organization (WHO) [1]. Lung cancer is estimated to affect about 2.21 million people worldwide and is the first major cause of cancer death

(1.80 million deaths). According to estimates, smoking cigarettes explains almost 70 to 80% of women's and nearly 90% of men's chance of lung cancer [2]. According to predictions, the death rate from cancer may rise to 60% until 2035 [3]. Malignant tumor prevalence has been observed to be increasing globally, which may be related to an increasing population. By histopathological type, malignancy can occur in any age group but is typically detected in the elderly age group of 50–60 years [4]. When lung cells mutate to grow uncontrollably, they aggregate into a cluster known as a tumor.

One of the types of lung cancer is squamous cell carcinoma. It happens when aberrant lung cells proliferate uncontrollably and form a tumor. The lymph nodes in and around the lungs, liver, bones, adrenal glands, and brain are places where tumor cells can spread (metastasize). Small-cell lung cancer and non-small cell lung cancer are the two main categories of lung cancer. Under a microscope, the cancerous cells of each type appear differently. They receive various treatments as well. Small-cell lung cancer has a worse outcome than non-small-cell lung cancer because the latter is more likely to be contained in one area. Therefore, the treatment would be more likely to be successful. Squamous cell carcinoma typically begins in the middle of the lungs. In comparison to tumors on the edges of the lungs, such as adenocarcinomas, these tumors may exhibit some symptoms earlier, such as coughing up blood. Because flow of fluids (blood and lymph) is constantly flowing through the lungs, squamous cell carcinoma frequently spreads (metastasizes) to other parts of the body. The fluids have the potential to spread cancer cells to adjacent tissues, including the esophagus, neck, chest wall, and sac that surrounds the heart. If it is not identified and treated promptly, it frequently spreads throughout the body [17].

Numerous causes, including exposure to breathing dangerous or toxic substances and a rise in the proportion of elderly individuals in society, have been linked to the growth in lung cancer cases around the world. However, it is more difficult to cure because the symptoms are unlikely to appear until the disease has progressed to other bodily areas. Although lung cancer can develop in people who have never smoked, those who do generally have an increased likelihood. The most frequent lung cancers are adenocarcinoma and squamous cell carcinoma, while other histological types include small as well as large cell carcinomas. Lung cancer adenocarcinoma primarily affects or has affected smokers, but it can also affect nonsmokers. This is more likely to affect women and younger people, and it first appears in the outer portions of the lungs before spreading. Smoking history is also linked to squamous cell carcinomas. On the other hand, small and large cell carcinomas can appear anywhere in the lung and have a propensity to spread and unroll quickly, making treatment more challenging [5].

One deep learning algorithm that can be utilized to categorize an image is CNN [18]. CNN is frequently used in image processing because it recognizes feature representations well. Deep learning has been applied in several biomedical fields with good accuracy. Deep Convolutional Neural Networks (DCNN), which are the basis for this accomplishment, are based on extracting features from the data itself in several layers [19, 20]. Convolutional Neural Networks (CNN) have made significant progress in the classification of many diseases, including brain diseases [21], the detection of breast cancer [22], skin cancer [23], arrhythmia detection [24], the detection of pulmonary pneumonia in X-ray images [25], the segmentation of fundus images [26], and the segmentation of the lung [27].

The goal of this study article is to test and analyze the proposed model for the classification of lung cancer using a Convolutional Neural Network (CNN) architecture. The paper is structured as follows in the following sections: The information in Sect. 2 describes previously explored research in the current domain. The used dataset is briefly introduced in Sect. 3. Section 4 goes into more detail about the proposed deep learning approach using CNN. All of the experimental observations and findings are reported in Sect. 5. Section 6 concludes the experiment while offering some suggestions for further work.

2 Related Work

Cancer is one of the major risks to humankind and is expected to be the top cause of mortality in the next few decades [7]. The analysis of clinical data was done using computer-aided diagnosis (CAD) to improve the efficiency and speed of cancer diagnosis. The field of CAD has undergone significant growth, and many machine learning techniques have been created for diagnosis purposes. Deep neural networks have demonstrated enhanced results in the detection of medical images among all methods of machine learning. Different CNN algorithms are applied to the classification of lung cancer images to improve the accuracy of the detection and classification. These precise predictions assist doctors by reducing their effort and avoiding diagnosis errors caused by human error.

By utilizing several data preprocessing strategies, Deep Convolutional Neural Network (DCNN) models for the classification of lung cancer images have shown improvements in accuracy and reduced model overfitting by using various data augmentation techniques [8]. For the classification of lung and colon cancer images on the LC25000 dataset, Satvik Garg and Somya Garg [9] created eight pre-trained CNN models with different feature extraction tools such as VGG16, InceptionV3, ResNet50, etc. They attained accuracy ranging from 96% to 100%. For the classification of lung cancer images, Bijaya Kumar and Himadri Chand [10] built a CNN model using cross-entropy as an error function and achieved training and validation accuracy of 96.11% and 97.2%, respectively.

Tri-category categorization of images of lung cancer by Baranwal et al. [11] using ResNet 50, VGG-19, Inception ResNet V2, and DenseNet to extract features, guide the CNN to improve inter-cluster closeness and decrease intra-cluster distance. In this work, Inception-ResNetv2 demonstrated a very high-test average accuracy of 99.7% in contrast to other models. When the triplet neural network model was trained using these four pre-trained models, DenseNet outperformed the other four with an accuracy level of 99.08%. Mangal et al. [12]’s classification performance of 97.8% was attained after analyzing digital pathology images of the colon and lung. A Convolutional Neural Network (CNN) with gamma correction was used by Setiawan et al. [13]. Gamma values of 0.8, 1.0, and 1.2 were employed and achieved the highest accuracy of 87.16% with a gamma value of 1.2. According to Manaswini et al. [13], a novel model is implemented to automatically classify the LC25000 lung histological image dataset. The EGOA-random forest model achieved an accuracy rate of 98.50%. A hybrid ensemble feature extraction approach was established by Talukder et al. [28] to effectively diagnose lung and colon cancer. The model has accuracy rates of 99.05%, 100%, and 99.30% for the detection

of lung, colon, and (lung and colon) cancer, respectively. While the proposed model has achieved average accuracy ranging from 99.3% to 99.8% in classifying lung malignant from benign lesions. The proposed model achieved the highest accuracies, compared by the state-of-the-art models.

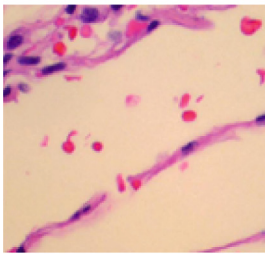
3 Used Dataset

The LC25000 Lung and Colon Histopathological Image Dataset contains 5000 images in each of the five classifications of colon adenocarcinoma, benign colonic tissue, lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue. The information has been verified and complies with HIPAA [14].

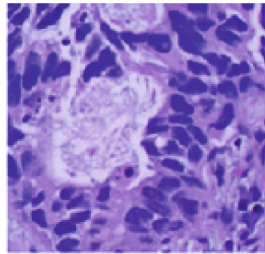
The maximum count of original images collected is only 750 total images of lung tissue and 500 total images of colon tissue, with each category receiving 250 images with a 1024×768 -pixel resolution. Python is used to reduce these images to 768×768 pixels, and then the augmenter software suite is used to expand them.

Therefore, there are 5000 images in each category of the larger and colon dataset. The `colon_image_sets`' subfolders contain `colon_aca` subfolder with 5,000 images of colon adenocarcinomas and `colon_n` subfolder with 5,000 images of benign colonic tissues. The `lung_image_sets`' subfolders contain three secondary subfolders: `lung_aca` subfolder with 5,000 images of lung adenocarcinomas, `lung_scc` subfolder with 5,000 images of lung squamous cell carcinomas, and `lung_n` subfolder with 5,000 images of benign lung tissues.

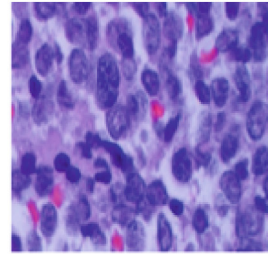
The left and right rotations, along with the horizontal and vertical flips, are used for augmentation [15]. The sample images for each category are shown in Fig. 1.



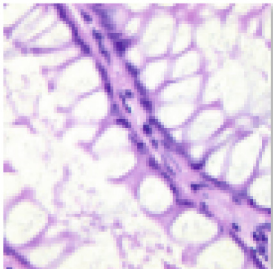
a) lung benign



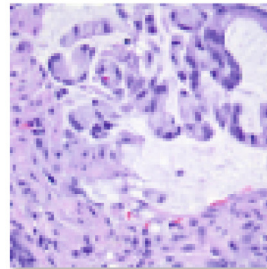
b) lung adenocarcinomas



c) lung squamous cell carcinomas



d) colon benign



e) colon adenocarcinoma

Fig. 1. a) Represents the benign histopathology of lung. b) and c) is an example image of adenocarcinomas and squamous cell carcinomas cancer types for the lung. d) Represents the benign histopathology of colon, and e) is an example image of adenocarcinomas of colon.

4 Proposed Deep Learning Approach Using CNN

This section discusses the proposed model for a lung cancer histopathology image classification. The dataset is divided randomly into two parts, training set and testing set. Additionally, the architecture of CNN [6] has been used in the proposed model, after making some modifications to achieve high performance in the classification.

The result of the training process is a model used for data testing. The output shows one of the two lung cancer classes as benign or malignant (squamous cell carcinomas) lung cancer.

Figure 2 shows the proposed architecture system. The architecture system consists of the following parts:

1. The dataset includes training data, which has been used for training the model.
2. Testing data is classified, after the training phase has been completed.
3. The CNN's architecture is an algorithm for feature extraction and classification.
4. The parameters of the trained model, that has been obtained in the training phase, is used in classification.
5. The lung cancer histopathology image classification is the output predicted by the model. There are two classes: benign and malignant (squamous cell carcinomas).

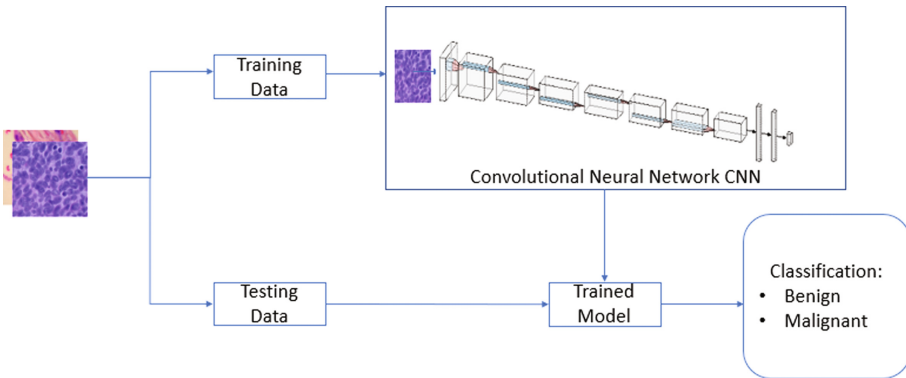


Fig. 2. Proposed architecture system

The used CNN's architecture is constructed using a stack of layers that has been used for recognition and classification of image. The training and testing data is passed through parameters such as max-pooling and kernel filters, before passing through the fully connected layer. It uses activation function ReLU in all hidden layers. A Softmax function is applied to classify the images with the following layers and parameters:

Input Layer. This layer has been used to load data and feed it to the first convolution layer. In this case the input is an image of size 224×224 pixels with color channels, which is 3 for RGB.

Convolution Layer. This layer has been used to convolve the input image with trainable filters to learn the geo spatial structure of images. This model contains six Convolution

Layers (CL) with a number of kernels equals to 64, 128, 256, 256, 512 and 512 respectively. The first CL layer that has been used is of kernel size (11×11). The stride is set to 4. Also, the padding has been set to valid. The other CL layers have (3×3) kernel size. The stride of those layers is set to 2, and the padding has been kept the same. In addition to this, ReLU activation has been applied for nonlinear operation to improve the performance.

Pooling Layer. Pooling operation has been used for down sampling the output images received from the convolution layer. After the second and fourth CL layers, there are two max pooling layers with (2×2) kernel size used. All the pooling layers have used the most common max pooling operation.

Flatten Layer. In order to connect a dense layer or fully connected layer, this layer has been used to convert the output from the convolution layer into a 1D tensor.

Fully Connected Layer or Dense Layer. These layers treat a simple vector as an input and produce an output in the form of vector. Three dense layers have been used in this model. The first, and the second dense layers contain 1024 neurons, and 512 neurons, respectively. The last one contains 2 neurons according to lung cancer classes. Output from the last fully connected layer could be activated with the help of Softmax activation.

Dropout Layer. A dropout layer with rate 0.4 in between fully connected layers, that randomly drops neurons from both visible and hidden layers, has been used to prevent overfitting of the model layers.

The CNN's setup is shown in Table 1, also Fig. 3. Describes the CNN architectures.

Table 1. Setup of CNN's model

Layer	#Filters/neurons	Filter size	Stride	#Nodes	Padding	Activation function
Conv 1	64	11×11	4×4	–	Valid	ReLU
Conv 2	128	3×3	1×1	–	–	ReLU
Max Pool1	–	2×2	–	–	–	–
Conv 3	256	3×3	1×1	–	–	ReLU
Conv 4	256	3×3	1×1	–	–	ReLU
Max Pool2	–	2×2	–	–	–	–
Conv 5	512	3×3	1×1	–	–	ReLU
Conv 6	512	3×3	1×1	–	–	ReLU
FC 1	–	–	–	1024	–	ReLU
FC 2	–	–	–	512	–	ReLU
Dropout	Rate = 0.4	–	–	–	–	–
FC3 (output)	–	–	–	2	–	Softmax

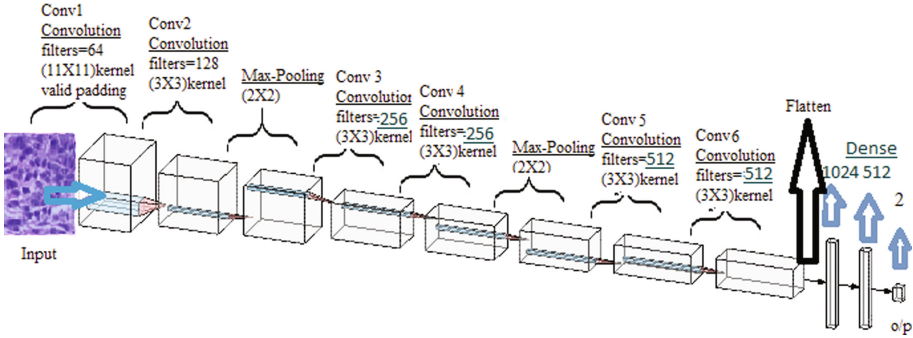


Fig. 3. The proposed CNN architecture.

5 Achieved Results

In this study the proposed model using CNN approach has been applied on LC25000 Lung histopathological images dataset using two classes of benign tissue, and malignant cells (squamous cell carcinoma) of lungs with 5000 histopathology images in each category. The model has been tried on four experiments using Google Colab. Each experiment with different percentage of data for training set and testing set. Also, each experiment has been applied four times with random selection of data for each training data and testing data. It has been applied to the same model with the same numbers of convolution layers with maximum number of epochs 50.

First Experiment (Exp. 1). In the Exp. 1, the used dataset has been split into 80% for training set and 20% for testing set. It has been applied four times with random selection of data. The proposed model contains three steps; features extraction, image classification then testing. Table 2 shows the accuracy of the Exp. 1 with 99.8% average accuracy.

Table 2. The accuracy results of First Experiment (Exp. 1)

Test no	Test loss (%)	Recognition accuracy (%)
1 st Test	1.67	99.06
2 nd Test	0.22	99.94
3 rd Test	0.00	100
4 th Test	0.16	100

Second Experiment (Exp. 2). In the Exp. 2, the used dataset has been split into 70% for training set and 30% for testing set. Also, it has been applied four times with random selection of data. The accuracy result of the Exp. 2 has been showed in Table 3. It has achieved average accuracy 99.8%.

Table 3. The accuracy results of First Experiment (Exp. 2)

Test no	Test loss (%)	Recognition accuracy (%)
1 st Test	0.09	100
2 nd Test	1.81	99.14
3 rd Test	0.18	99.93
4 th Test	0.07	100

Third Experiment (Exp. 3). In the Exp. 3, the used dataset has been split into 50% for training set and 50% for testing set. It has been applied four times with random selection of data. The model applied in Exp. 3 has achieved 99.5% average accuracy. Table 4 shows the result accuracy of the Exp. 3.

Table 4. The accuracy results of First Experiment (Exp. 3)

Test no	Test loss (%)	Recognition accuracy (%)
1 st Test	0.38	99.90
2 nd Test	2.07	99.20
3 rd Test	0.00	100
4 th Test	1.85	98.90

Fourth Experiment (Exp. 4). In the Exp. 4, the used dataset has been split into 40% for training set and 60% for testing set four times with random selection of data. It has achieved 99.3% average accuracy. Table 5 shows the result accuracy of the Exp. 4.

Table 5. The accuracy results of First Experiment (Exp. 4)

Test number	Test loss (%)	Recognition accuracy (%)
1 st Test	4.24	98.37
2 nd Test	0.47	99.87
3 rd Test	0.49	99.75
4 th Test	1.18	99.37

Figure 4 shows the comparison between the four tests has been done for each experiment according to the achieved accuracy. Each test has been applied on the proposed model of classification images. The model has been applied four times with random

selection of data for each experiment. It has achieved average accuracy 99.8%, 99.8%, 99.5%, and 99.3% for Exp. 1, Exp. 2, Exp. 3, and Exp. 4, respectively. The calculated standard deviation for each experiment is in the range from ± 0.14 to ± 0.28 .

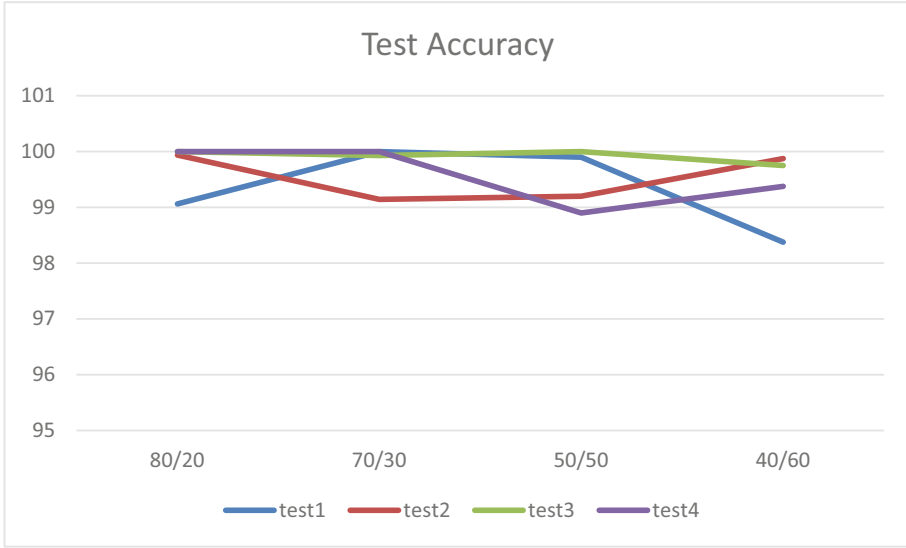


Fig. 4. The comparison between the four tests that has been done for each experiment according to the achieved accuracy

6 Conclusions

Lung cancer causes a very high mortality all over the world. Therefore, deep learning techniques are used to speed up the critical process of diagnosing lung cancer and reduce the workload of pathologists.

This research proposed a lung cancer classification model using CNN's architecture. The proposed model has been applied on LC25000 dataset of lung histopathological images. The proposed model has classified an image into two categories: benign and malignant cells. Each class of lung has 5000 histopathology images.

The proposed model has been applied in four experiments. First Experiment (Exp. 1), has 80% of the dataset goes into the training set, and 20% of the dataset goes into the testing set. Second Experiment (Exp. 2), has split dataset into 70% for the training set, and 30% for the testing set. Third Experiment (Exp. 3), has split dataset into 50% for the training set, and 50% for the testing set. Fourth Experiment (Exp. 4), has split dataset into 40% for the training set, and 60% for the testing set.

Each experiment has been applied four times with random selection of data. Using Google Colab, each experiment has used the same model with the same number of convolution layers and a maximum number of epochs 50 to measure the model performance. The model has achieved high performance compared with the state-of-the-art models.

The proposed model has achieved average accuracy of 99.8%, 99.8%, 99.5%, and 99.3% for Exp. 1, Exp. 2, Exp. 3, and Exp. 4, respectively. The results, compared by the state-of-the-art models, achieved the highest accuracies.

It was experimentally proved that, there is no tangible difference in average accuracy between the applied experiments. Therefore, the proposed model has been efficient in comparison with the state-of-the-art. In future work, different CNN architectures can be explored. Also, other datasets can be used to optimize the hyperparameters.

References

1. World Health Organization: Emergencies preparedness, response, disease outbreak news World Health Organization (WHO) (2020). Pneumonia of unknown cause—China
2. Walser, T., et al.: Smoking and lung cancer: the role of inflammation. *Proc. Am. Thorac. Soc.* **5**(8), 811–815 (2008)
3. Araghi, M., et al.: Global trends in colorectal cancer mortality: projections to the year 2035. *Int. J. Cancer* **144**(12), 2992–3000 (2019)
4. Arslan, N., et al.: Analysis of cancer cases from Dicle University Hospital; ten years' experience. *J. Clin. Anal. Med.* **9**(2), 102–106 (2018)
5. American Cancer Society. <https://www.cancer.org/cancer/lung-cancer/causes-risks-prevention/risk-factors.html>
6. Ghaleb, M.S., Ebied, H.M., Shedeed, H.A., Tolba, M.F.: Image Retrieval based on deep learning. *J. Syst. Manag. Sci.* **12**(2), 477–496 (2022)
7. Tang, J., et al.: Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Trans. Inf. Technol. Biomed.* **13**(2), 236–251 (2009)
8. Teramoto, A., et al.: Automated classification of lung cancer types from cytological images using deep convolutional neural networks. *BioMed Res. Int.* (2017)
9. Garg, S., Garg, S.: Prediction of lung and colon cancer through analysis of histopathological images by utilizing pre-trained CNN models with visualization of class activation and saliency maps. In: 2020 3rd Artificial Intelligence and Cloud Computing Conference, pp. 38–45 (2020)
10. Hatuwal, B.K., Thapa, H.C.: Lung cancer detection using convolutional neural network on histopathological images. *Int. J. Comput. Trends Technol.* **68**(10), 21–24 (2020)
11. Baranwal, N., Doravari, P., Kachhoria, R.: Classification of histopathology images of lung cancer using convolutional neural network (CNN). arXiv preprint [arXiv:2112.13553](https://arxiv.org/abs/2112.13553) (2021)
12. Mangal, S., Chaurasia, A., Khajanchi, A.: Convolution neural networks for diagnosing colon and lung cancer histopathological images. arXiv preprint [arXiv:2009.03878](https://arxiv.org/abs/2009.03878) (2020)
13. Setiawan, W., et al.: Histopathology of lung cancer classification using convolutional neural network with gamma correction. *Commun. Math. Biol. Neurosci.* (2022). Article ID 81
14. Pradhan, M., Sahu, R.K.: Automatic detection of lung cancer using the potential of artificial intelligence (AI). In: *Machine Learning and AI Techniques in Interactive Medical Image Analysis*, pp. 106–123. IGI Global (2023)
15. Borkowski, A.A., et al.: Lung and colon cancer histopathological image dataset (lc25000). arXiv preprint [arXiv:1912.12142](https://arxiv.org/abs/1912.12142) (2019)
16. Šarić, M., et al.: CNN-based method for lung cancer detection in whole slide histopathology images. In: 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech), pp. 1–4. IEEE (2019)
17. Squamous Cell Carcinoma of the Lung - Harvard Health. <https://www.health.harvard.edu/cancer/squamous-cell-carcinoma-of-the-lung>

18. Albawi, S., Mohammed, T.A., Al-Zawi, S.: Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET), pp. 1–6. IEEE (2017)
19. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
20. Jiang, X.: Feature extraction for image recognition and computer vision. In: 2009 2nd IEEE International Conference on Computer Science and Information Technology, pp. 1–15. IEEE (2009)
21. Talo, M., et al.: Convolutional neural networks for multi-class brain disease detection using MRI images. *Comput. Med. Imaging Graph.* **78**, 101673 (2019)
22. Celik, Y., et al.: Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. *Pattern Recognit. Lett.* **133**, 232–239 (2020)
23. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**(7639), 115–118 (2017)
24. Yoon, S.H., et al.: Chest radiographic and CT findings of the 2019 novel coronavirus disease (COVID-19): analysis of nine patients treated in Korea. *Korean J. Radiol.* **21**(4), 494–500 (2020)
25. Rajpurkar, P., et al.: Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint [arXiv:1711.05225](https://arxiv.org/abs/1711.05225) (2017)
26. Tan, J.H., et al.: Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. *Inf. Sci.* **420**, 66–76 (2017)
27. Gaál, G., Maga, B., Lukács, A.: Attention u-net based adversarial architectures for chest X-ray lung segmentation. arXiv preprint [arXiv:2003.10304](https://arxiv.org/abs/2003.10304) (2020)
28. Talukder, Md.A., et al.: Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. *Expert Syst. Appl.*, 117695 (2022)



An Enhanced Deep Learning Approach for Breast Cancer Detection in Histopathology Images

Mahmoud Ouf^{1(✉)}, Yasser Abdul-Hamid¹, and Ammar Mohammed^{1,2}

¹ Department of Computer Science Faculty of Graduate Studies for Statistical
Research Cairo University, Cairo, Egypt

mahmoudoof84@gmail.com, {yabdelfattah, ammar}@cu.edu.eg

² Faculty of computer Science, October University for Modern Science and Arts,
6th of October, Egypt

Abstract. Breast cancer is defined as abnormal cellular proliferation in the breast. The most common kind of cancer that affects the breast and causes mortality in women is invasive ductal carcinoma (IDC). As a result, early diagnosis and prognosis have become critical to maximize survival and minimize mortality. Mammograms, computerized tomography (CT) scans, and ultrasounds are among the breast cancer tests available. On the other hand, histopathology evaluation with a biopsy is regarded as one of the most trustworthy techniques for determining if suspicious lesions are malignant. This paper proposes an enhanced approach for classifying breast tumors using a proposed CNN model that we named (Cancer-Net). We evaluate the proposed model on a benchmark dataset containing 277,524 patches. Compared to several types of CNN-based models, our proposed model has achieved accuracy, Area Under Curve (AUC), precision, recall, and F1-score of 86%, 92%, 81%, 84%, and 83%, respectively, outperforming the previous work on the same benchmark.

Keywords: Deep learning · Convolutional neural network · Depthwise separable network · Cancer classification · Invasive ductal carcinoma (IDC)

1 Introduction

Our bodies have millions of cells. Cancer starts when cellular changes cause these cells to grow excessively, creating a lump called a primary tumor [1]. According to Changing profiles of cancer burden worldwide and in China, a secondary analysis of the global cancer statistics in 2020, It became clear that among the 14 most common cancers, deaths due to breast cancer increased the most [2–4]. Manual diagnosis of breast cancer is tedious and time-consuming; thus, there arises a need for automatic diagnosis. The current healthcare systems have proven to be helpful, but they are prone to errors. Therefore, medical image classification using Computer-Aided Diagnosis (CAD) has emerged as an efficient tool that can help doctors classify medical images into different categories, leading to early diagnosis and treatment [5–7]. In this context, machine learning and deep

learning have developed algorithms capable of more accurately diagnosing the disease at an earlier stage, thus AI techniques can boost the procurement of new fidelity protocols in medicine and reduce healthcare costs due to misdiagnosis [8]. Also, Manual classification for breast cancer is not effective enough. There are three main reasons: (1) the professional background and rich experience of pathologists and radiologists are so difficult to inherit or innovate that primary-level hospitals and clinics suffer from the absence of skilled pathologists, (2) the tedious task is expensive and time-consuming, (3) over fatigue of pathologists might lead to misdiagnosis. Hence, it is extremely urgent and important to use computer-aided breast cancer classification, which can reduce pathologists' heavy workloads and help avoid misdiagnosis. The decision of an optimal therapeutic schedule for breast cancer rests upon refined classification. One main reason is that doctors who know the subordinate classes of breast cancer can control the metastasis of tumor cells early and make substantial therapeutic schedules according to special clinical performance and prognosis results of multiple breast cancers.

Many different types of breast cancer images are diagnosed through deep-learning techniques [4, 9, 10]. These types include Mammography, Magnetic resonance, Ultrasounds, and Histopathology images. Histopathology images are microscopy images of Nuclei patches and cytoplasm that appear purple and pinkish due to the hematoxylin and eosin staining. The diagnosis using histopathology images is a gold standard for identifying breast cancer compared with other medical imaging as it is more reliable [11, 12].

This paper proposes an enhanced deep learning model for classifying breast cancer from histopathological images we named (CancerNet) and compares the experimental results to several pre-trained models by transfer learning such as VGG16, VGG19, and ResNet50 on the same dataset. The proposed model is trained on a benchmark breast cancer images invasive ductal carcinoma (IDC) dataset and evaluated under several performance metrics [13].

The adopted approach is extending the work done previously using depth-wise separable convolution Network (MobileNet and Xception) architecture to improve performance. Also, we evaluating the proposed model based on different metrics such as accuracy, precision, recall, f1-score, and Area Under Curve (AUC). By these types of metrics, we can have a clear indicator of the generalization of the model over the imbalanced histopathology images dataset. The experimental result shows that the proposed deep learning model is competitive compared to the other approaches.

The rest of this paper is organized as follows: Sect. 2 describes some related research efforts. Section 3 describes the proposed model. Section 4 introduces the experimental results and the model evaluation and compares our results with the other published work. Section 5 finally concludes the paper.

2 Related Work

This section presents the description and method for extracting histopathological images as well as the contributions of researchers in recent years regarding the diagnosis of breast cancer through deep learning techniques.

2.1 Histopathology Images Description

As we mentioned early histopathology images are a gold standard for identifying the breast cancer compared with other medical imaging, e.g., mammography, magnetic resonance (MR), and computed tomography (CT) as it is more reliable. It basically is a microscopy image patches Nuclei, and cytoplasm appear purple and pinkish, respectively, due to the hematoxylin and eosin staining as shown below in Fig. 1 [11].

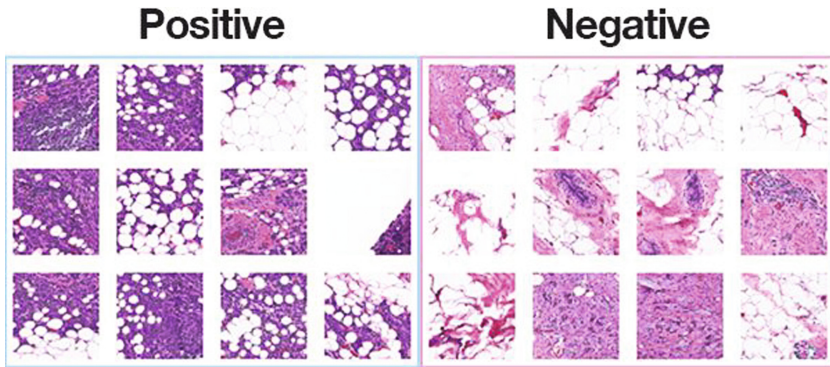


Fig. 1. Examples of both positive and negative samples of Histopathology Images [14].

Needle cell aspiration is performed under local anesthesia in the breast area, by injecting the drug into the skin around the desired area. Needle prick may cause slight pain. After that, the doctor inserts a long needle into the breast tissue. The pathologist begins with a macroscopic examination. The fragment of tissue is then cut into ultrathin slices which are placed on glass slides and coloured using different chemical products. The pathologist analyses these slides under the microscope and makes a diagnosis.

2.2 Previous Research

In recent years, many researchers have used the histopathology datasets to categorize images due to the widespread availability of a large set of images. In [15, 16], the authors used the Break-His dataset and have attained significant performance on classifying breast cancer using histopathology image by comparing

the results separately on patch-wise and image-wise classification. In addition to data augmentation, transfer learning can also be utilized when the dataset is not enough for diagnosis. Li Y et al. and Deniz E et al. discussed histology image classification using several pre-trained networks such as Res-Net50, Alex-Net, VGG16, and obtained satisfactory results [17,18]. In another research, Brancati et al. detected invasive ductal carcinoma. The researchers also performed multi-classification of lymphoma in histological images, using residual convolutional auto-encoder (CAE) and implemented two approaches. Firstly, classification by reconstruction, training CAE in an unsupervised manner, and secondly, supervised classification by guiding only the encoder part of Fusion-Net called Supervised Encoder Fusion-Net (SEF). They concluded that SEF provides outstanding results for the considered cases and can be used in the future for other histopathology image analysis cases [19].

Jiang et al. proposed a new model, Breast Cancer Histopathology Image Classification network (BHC Net). The authors reduced training parameters by designing a new module called the small Squeeze and Excitation Residual network (SE-ResNet) and prevented the model from the risk of overfitting. The proposed method attained an accuracy of 98.87 to 99.34% for binary classification and 90.66 to 93.81% for multi-class classification [20]. Nahid et al. proposed three models for histopathology image classification CNN, long-short-term-memory (LSTM), and a combination of CNN and LSTM. The authors found unseen structured and statistical information from the data using two unsupervised clustering techniques-K-Means and Mean shift. CNN with Softmax classifier outperforms the other two models with an accuracy of 91% [21]. The authors in [22] presented a hybrid deep learning approach, multi-modal data fusion by combining structured data from Electronic Medical Record (EMR) with pathological images. Results show significant improvement in diagnostic accuracy by fusing high dimensional and low dimensional data. Sharma and Mehra suggested a framework for the multi-classification of histopathology images on the Break His dataset. The authors, firstly, employed traditional classifiers for performing classification using hand-crafted features. Secondly, transfer learning was implemented, where pre-trained models “VGG16,” “VGG19,” and “ResNet50” were used as feature extractors and as baseline models. Outcomes show that the approach based on pre-trained models as a feature extractor outperforms the baseline and hand-crafted approach. VGG16 model with a linear SVM classifier attained the highest accuracy [23].

Similar to the work of this paper, several authors have used the same benchmark IDC dataset to obtain a good breast tumor classifier. The data was first used by the authors in [13] system adapts a 3-layers CNN architecture employing 16, 32, and 128 neurons, achieving 71.8% (F-measure) and 84.23% (BAC). Precision, recall with 65.4%, 79.6% respectively. In [24], the authors trained the whole slid images (WSI) using the proposed CNN model. 27 hidden layers are connected. These include dense architecture, max pooling, batch normalization, dropout, and activation. During training, model achieved up to 83% accuracy with a loss of 0.39 in the same dataset. In another work [25], the authors have

down-sampled their original $\times 40$ images by a factor of 16:1, for an apparent magnification of $\times 2.5$. They attempted three different approaches of using these (50*50) patches and casting them into the (32 * 32) solution domain and the model achieved accuracy 84% and an f1-score of 76%.

3 Proposed Approach

This section describes the design of the proposed model CancerNet as shown in Fig. 2. We designed a network appropriately which: (1) Uses exclusively 3×3 CONV filters, similar to VGG Net. (2) Stacks multiple 3×3 CONV filters on top of each other prior to performing max-pooling. (3) Uses depth-wise separable convolution rather than standard convolution layers. Depthwise separable convolution is not a new idea in deep learning. In fact, the first utilized to depth-wise network by Google Brain interns. (a) It is more efficient. (b) Requires less memory. (c) Requires less computation. (d) Can perform better than standard convolution in some situations [26].

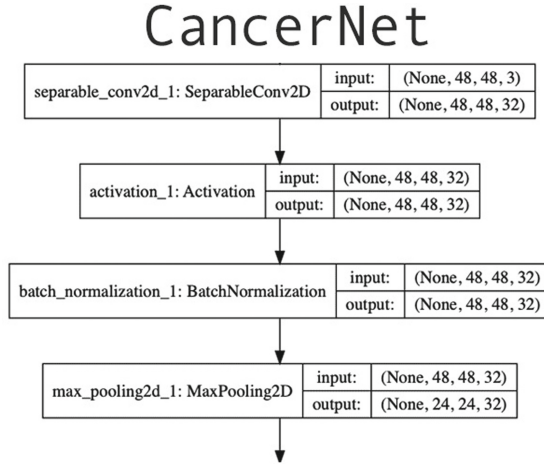


Fig. 2. Our deep learning classification architecture for predicting breast cancer.

Three blocks of (SeparableConv2D, Activation = RELU, MaxPooling2D) are defined with increasing stacking and number of filters, applied BatchNormalization and Dropout as well to avoid overfitting. Finally, a fully connected (Flatten, Dense) is added as a head layer. Our fully connected and RELU layers and softmax classifier make the head of the network. The output of the softmax classifier will be the prediction percentages for each class our model will predict. We used Data augmentation, such as horizontal and vertical flips, shifting, zooming and shearing a form of regularization, that is important for nearly all deep-learning experiments to assist with model generalization. It allows to reduce overfitting

and increases generalization performance by enhancing the training dataset itself which is the root of the problem, without relying on doing any changes to the model architecture. Furthermore, Data augmentation is a method of intentionally changing training models and their appearance slightly before passing them on to the network for training. This partially reduces the need to collect more training data.

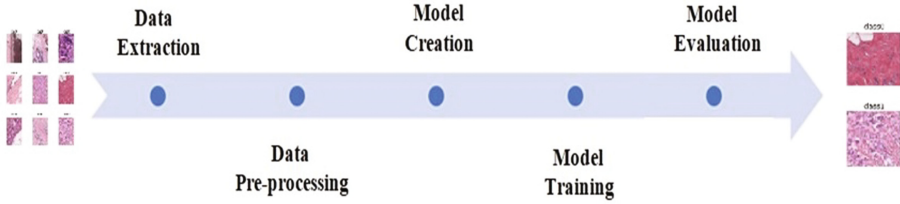


Fig. 3. Proposed approach pipeline.

We have proposed a pipeline for creating our deep learning model CancerNet. The pipeline is composed of multiple stages, from receiving the raw data to producing the model classification output. These stages include data extraction, data preprocessing, model Creation, model training and evaluation. Each stage output in the pipeline is given as input to the subsequent stage. The pipeline of the proposed approach is shown in Fig. 3. Data defines the task and plays a big part in the model performance. After obtaining the breast cancer dataset, we applied the required conversions and preprocessing of the images to help the model perform at its best.

4 Experimental Results

This section presents the dataset’s experimental classification results on histopathological test images. Before showing the experimental results, we will describe the dataset that used in our case.

4.1 Dataset Description

We use Invasive Ductal Carcinoma (IDC) dataset, the most common breast cancer dataset. The dataset was originally curated by Janowczyk, Madabhushi, and Roa et al. but is available in the public domain on the Kaggle website [27]. The original dataset consisted of 162 slide images scanned at $40\times$. Slide images are naturally massive (in terms of spatial dimensions), so to make them easier to work with, a total of 277,524 patches of 50×50 pixels were extracted, including 198,738 negative examples (no breast cancer), 78,786 positive examples (indicating breast cancer was found in the patch). There is an imbalance in the class data with over $2\times$ the number of negative data points than positive data points.

4.2 Results

Our proposed model CancerNet was trained using Google Colaboratory resources with Jupyter Notebooks environment from Google, NVIDIA Tesla T4 GPU, Intel(R) Xeon(R) CPU @ 2.20 GHz, and 12.68 GB RAM. The model was implemented on Python 3.7.15, using Keras 2.6 library with TensorFlow 2.9.2 as a backend. It took an average of 430s to complete each epoch. The model results are discussed and evaluated in this section.

Table 1. comparison Results of experiments using different models on IDC dataset.

Performance measure	VGG16	VGG19	ResNet50	Cancer Net
Precision (benign)	78%	95%	85%	93%
Recall (benign)	98%	69%	86%	86%
Precision (malignant)	87%	54%	64%	70%
Recall (malignant)	29%	91%	62%	83%
F1-score (benign)	87%	68%	86%	89%
F1-score (malignant)	44%	80%	63%	76%
Accuracy	79%	75%	79%	86%
Sensitivity	98%	69%	86%	86%
Specificity	30%	91%	62%	83%
Precision (Macro-avg)	83%	74%	75%	81%
Recall (Macro-avg)	64%	80%	74%	84%
F1-score (Macro-avg)	65%	74%	74%	83%
AUC	64%	80%	74%	92%

We split the cancer image dataset IDC into training, validation, and testing with split ratios of 0.7, 0.1, and 0.2, respectively. Table 1 shows the overall performance metrics results. Those results show the comparison between Transfer learning models using VGG16, VGG19, and ResNet50 fine-tuning and Our model (CancerNet), which uses depthwise separable convolution architecture. While this 3.10 GB deep learning dataset is not large compared to most datasets, we will use the opportunity to put the (Keras-ImageDataGenerator) to work, yielding small batches of images. This eliminates the need to have the whole dataset in memory. We have used Depthwise separable convolution Network architecture as the baseline for our CNN model. Since the medical image datasets are imbalanced, we applied different augmentation techniques to maximize our data during run-time. The applied augmentation approaches are: rotation range = 20, re-scale = 1/255.0, zoom range = 0.05, height shift range = 0.1, width shift range = .1, horizontal and vertical flip = True, shear range = 0.5, and fill mode = nearest. Optimized hyper-parameters used for training our proposed model are: (1) Optimizer = Adagrad, (2) Number of Epochs = 30, (3) Batch size = 32, (4) Learning rate = 0.01, (5) Learning rate decay = 0.0004.

The output of our classification model was analyzed and evaluated using several evaluation metrics. Usually, most of the previous studies used accuracy as a metric to evaluate the model's performance. However, using this measure alone for comparison can be misleading, as it ignores the sensitivity to imbalanced data, and the performance of some classes becomes better than others. We have used other measures such as precision, recall, and f1-score as shown in Eq. 1, 2 to have a clear indicator of the model's generalization over the imbalanced histopathological images dataset. Macro-average is used for overall computing performance by calculating the results for each class separately and then getting the average value. Our CNN model yields the best overall performance regarding the recall, Area under Curve (AUC), precision, F1-score and accuracy (0.84, 0.92, 0.81, 0.83, and 0.86), respectively, outperforming Transfer learning models and previous work in [13, 24, 25, 28] as shown in Table 2.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP}, F1Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN}, Specificity = \frac{TN}{FP + TN} \quad (3)$$

To understand our model's performance at a deeper level, we compute the sensitivity and the specificity as shown in Eq. 3. Our sensitivity measures the proportion of the true positives also predicted as positive (86%). Conversely, specificity measures our true negatives (83%). There is always a balance between sensitivity and specificity that a machine learning/deep learning engineer must manage. However, that balance becomes extremely important when it comes to deep learning and healthcare treatment.

Table 2. Previous works for breast tumor classification using the same benchmark.

Author	Year	Performance
Cruz et al. [13]	2014	Accuracy: 84%, precision: 65%, recall: 79%, F1-score: 71%
Janowczyk et al. [25]	2016	Accuracy: 84%, F1-score: 76%
Abdolahi et al. [28]	2020	Accuracy: 85%, F1-score: 83%
K. Kumar et al. [24]	2021	Accuracy: 83%
This work	2022	Accuracy:86%,precision:81%, recall:84%,F1:83%

5 Conclusion

Breast cancer is one of the most causes of death in women. Early diagnosis and detection of Invasive Ductal Carcinoma (IDC) are important for the treatment. In this paper, we have proposed an enhanced deep-learning model for breast cancer classification from histopathology images. According to the results,

our baseline models yielded significant improvement in comparison with the transfer learning methods for pre-trained networks such as VGG16, VGG19, and ResNet50 on this dataset. We have performed various data augmentation techniques, increasing the dataset size and improving accuracy. The model was trained on a benchmark (IDC) of 277,524 images and was evaluated under several performance metrics such as precision, recall, f1-score, Area Under Curve (AUC) and accuracy. We have achieved accuracy and AUC of “ 86%, 92% ” respectively, a state-of-the-art result compared to previous literature. For future work, we intend to extend this work to experiment with more cancer types.

References

1. Goyal, K., Sodhi, P., Aggarwal, P., Kumar, M.: Comparative analysis of machine learning algorithms for breast cancer prognosis. In: Krishna, C.R., Dutta, M., Kumar, R. (eds.) *Proceedings of 2nd International Conference on Communication, Computing and Networking. LNNS*, vol. 46, pp. 727–734. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1217-5_73
2. Cao, W., Chen, H.-D., Yu, Y.-W., Li, N., Chen, W.-Q.: Changing profiles of cancer burden worldwide and in china: a secondary analysis of the global cancer statistics 2020. *Chin. Med. J.* **134**(07), 783–791 (2021)
3. Globocan 2018: India factsheet—cancerindia.org.in. <http://cancerindia.org.in/globocan-2018-india-factsheet/>
4. Chugh, G., Kumar, S., Singh, N.: Survey on machine learning and deep learning applications in breast cancer diagnosis. *Cogn. Comput.* **13**(6), 1451–1470 (2021)
5. Khuriwal, N., Mishra, N.: Breast cancer diagnosis using deep learning algorithm. In: *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 98–103. IEEE (2018)
6. Selvathi, D., Aarthi Poornila, A.: Deep learning techniques for breast cancer detection using medical image analysis. In: Hemanth, J., Balas, V.E. (eds.) *Biologically Rationalized Computing Techniques For Image Processing Applications. LNCVB*, vol. 25, pp. 159–186. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-61316-1_8
7. Lai, Z., Deng, H.: Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron. *Comput. Intell. Neurosci.* **2018** (2018)
8. Duggento, A., et al.: A random initialization deep neural network for discriminating malignant breast cancer lesions. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 912–915. IEEE (2019)
9. Alkhouly, A.A., Mohammed, A., Hefny, H.A.: Improving the performance of deep neural networks using two proposed activation functions. *IEEE Access* **9**, 82249–82271 (2021)
10. Mohammed, A., Kora, R.: An effective ensemble deep learning framework for text classification. *J. King Saud Univ.-Comput. Inf. Sci.* **34**(10), 8825–8837 (2022)
11. Rashmi, R., Prasad, K., Udupa, C.B.K.: Breast histopathological image analysis using image processing techniques for diagnostic puposes: a methodological review. *J. Med. Syst.* **46**(1), 1–24 (2022)

12. Eldin, S.N., Hamdy, J.K., Adnan, G.T., Hossam, M., Elmasry, N., Mohammed, A.: Deep learning approach for breast cancer diagnosis from microscopy biopsy images. In: 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), pp. 216–222 (2021)
13. Cruz-Roa, A., et al.: Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: Medical Imaging 2014: Digital Pathology, vol. 9041, p. 904103. SPIE (2014)
14. Rosebrock, A.: Breast cancer classification with Keras and Deep Learning - PyImageSearch—pyimagesearch.com. <https://pyimagesearch.com/2019/02/18/breast-cancer-classification-with-keras-and-deep-learning/>
15. Araújo, T., et al.: Classification of breast cancer histology images using convolutional neural networks. *PloS One* **12**(6), e0177544 (2017)
16. Roy, K., Banik, D., Bhattacharjee, D., Nasipuri, M.: Patch-based system for classification of breast histology images using deep learning. *Comput. Med. Imaging Graph.* **71**, 90–103 (2019)
17. Li, Y., Wu, J., Wu, Q.: Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning. *IEEE Access* **7**, 21400–21408 (2019)
18. Deniz, E., Şengür, A., Kadiroğlu, Z., Guo, Y., Bajaj, V., Budak, Ü.: Transfer learning based histopathologic image classification for breast cancer detection. *Health Inf. Sci. Syst.* **6**(1), 1–7 (2018). <https://doi.org/10.1007/s13755-018-0057-x>
19. Brancati, N., De Pietro, G., Frucci, M., Riccio, D.: A deep learning approach for breast invasive ductal carcinoma detection and lymphoma multi-classification in histological images. *IEEE Access* **7**, 44709–44720 (2019)
20. Jiang, Y., Chen, L., Zhang, H., Xiao, X.: Breast cancer histopathological image classification using convolutional neural networks with small se-resnet module. *PloS One* **14**(3), e0214587 (2019)
21. Nahid, A.-A., Mehrabi, M.A., Kong, Y.: Histopathological breast cancer image classification by deep neural network techniques guided by local clustering. *BioMed Res. Int.* **2018** (2018)
22. Yan, R., et al.: Integration of multimodal data for breast cancer classification using a hybrid deep learning method. In: Huang, D.-S., Bevilacqua, V., Premaratne, P. (eds.) ICIC 2019. LNCS, vol. 11643, pp. 460–469. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26763-6_44
23. Sharma, S., Mehra, R.: Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight. *J. Dig. Imaging* **33**(3), 632–654 (2020)
24. Kumar, K., Saeed, U., Rai, A., Islam, N., Shaikh, G.M., Qayoom, A.: Idc breast cancer detection using deep learning schemes. *Adv. Data Sci. Adapt. Anal.* **12**(02), 2041002 (2020)
25. Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *J. Pathol. Inf.* **7**(1), 29 (2016)
26. He, Y., Qian, J., Wang, J.: Depth-wise decomposition for accelerating separable convolutions in efficient convolutional neural networks. *arXiv preprint arXiv:1910.09455* (2019)
27. Breast Histopathology Images—kaggle.com. <https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images>
28. Abdolahi, M., Salehi, M., Shokatian, I., Reiazi, R.: Artificial intelligence in automatic classification of invasive ductal carcinoma breast cancer in digital pathology images. *Med. J. Islamic Rep. Iran* **34**, 140 (2020)



Reducing Deep Learning Complexity Toward a Fast and Efficient Classification of Traffic Signs

Btissam Bousarhane^(✉)  and Driss Bouzidi 

Smart Systems Laboratory (SSL), National School of Computer Science and Systems Analysis
ENSIAS, Mohammed V University, Rabat, Morocco
{ibtissam_bousarhane, driss.bouzidi}@um5.ac.ma

Abstract. In our digital society, where the technology is changing at an accelerated rate, many of human activities have become easier due to the exponential growth of many intelligent devices, especially mobiles ones. These technologies have changed many aspects of our daily life, including among others the way we use and control our cars. Improving the driving experience, and enhancing the environmental perception to support safety, especially in VANETs, represents in effect one of the most recent application fields of these technologies. This safety improvement includes among others traffic signs recognition. Although the high performances of 2D CNNs in ensuring this recognition, their computational complexity limits however their use in this type of devices characterized by their low resources. While we find recently that the 1D networks gains more attention and popularity in many areas of research, and that due to their low-cost implementation and computational complexity compared to the traditional 2D CNNs. Therefore, to benefit from the advantages of these two types of networks, we have applied instead of the full connections between the nodes, a final 1D module based essentially on one-dimensional separable convolutions to the final features' vector, and that using different sets of kernels. The training, the validation and the testing results show that, the proposed approach is fast and efficient in terms of inference time and classification accuracy, and that using low resources environments and well-known public datasets (Belgium and CURE-TSR datasets).

Keywords: Traffic signs recognition · Road signs classification · Deep learning · CNNs · BTSCD

1 Introduction

Recently, Deep Learning (DL) approaches have gained many attentions, and that due to their outstanding performances in different fields of research, including traffic signs recognition. This recognition plays in fact an important role in ameliorating the environmental perception in the context of VANETs (Vehicular Ad-Hoc Networks), and also in improving drivers' experiences with its implementation in ADAS (Advanced Driver Assistance Systems) and self-driving cars. While this type of approaches, including CNNs, gain considerably in terms of accuracy, even under adverse conditions, they fail however in ensuring the real-time recognition in real world-situations, because they are

very expensive and demanding when considering the computational load and hardware requirements [1]. To overcome this huge challenge, increasing CNNs' depth and size is no longer an option, especially when dealing with low and limited resources environments. Looking hence for other solutions that could outperform the traditional DL architectures, it is what is needed to achieve a fast and efficient detection, and classification of traffic signs.

Concerning the traditional CNNs' architecture, we find that the 1D as well as the 2D CNNs are in effect based on one-dimensional fully-connected layers in order to ensure the connection between the hidden nodes. Which by consequence leads to increase the computational complexity of these networks, especially when dealing with deep CNNs of larger size and depth. Thing that represents one of the main drawbacks of this type of networks. To reduce this complexity, dealing with this one-dimensionality using a different approach is the main objective of this paper. From this perspective, the paper is organized as follow: the second section presents some relevant works concerning traffic signs classification using DL. The third section shows the proposed approach, while the fourth section discusses the testing results. Finally, the conclusion and the perspectives are presented in the last section.

2 Related Works

Traffic signs recognition is a field of research that is widely treated in the literature [1]. Hence, many approaches are proposed by researchers in order to ensure this recognition. Although the fact that each proposed approach has its own advantages and limitations, Deep Learning [2] methods have proven however their superiority, by ensuring high and accurate recognitions under many real-world adverse conditions. Accordantly, due to their outstanding performances, they are adopted by many researchers for traffic signs recognition, especially for the classification stage using CNNs.

In this context, we can mention the work of Mehta, Paunwala and Vaidya [3]. The proposed network includes 3 convolutional layers, where each layer is followed by a maxpooling one, and finally 2 other fully-connected layers are added. To achieve better results, Dropout is added after each of the fully-connected ones. Furthermore, Relu is used as the activation function, in addition to Adam, and Softmax for the final predictions. To test the classification performances, Belgium Traffic Signs Recognition Dataset (BTSRD) is used. The trainable parameters of the network are almost 10 342 810, and the achieved accuracy for the testing set is about 97.06%, using Intel (R) Core (TM) i5-7500 CPU @ 3.40 GHz, 8 GB RAM.

Another approach based on CNNs and Adam optimizer is proposed by Suriya [4]. The work represents an extension of LeNet-5 CNN model. To reduce the effect of illumination changes, the detection stage is based on HSV color space, while the classification is based on the improved LeNet-5 CNN. The proposed network includes 2 convolutional layers, where each layer is also followed by a pooling one. In addition to that, two fully-connected layers are used before getting the final predictions. The adopted approach is evaluated using six categories of traffic signs. The obtained accuracy is about 99.70%, where each sign is recognized within 8 ms per image.

In addition to the illumination changes, other types of challenging conditions are considered by the work of Zhou, Zhan and Fu [5]. The severe conditions treated include

specifically ice and snow environment, including other real-world situations, like occlusions, lightening variations, etc. The adopted network is a Learning Region-Based Attention Network, designed for the detection and the classification of traffic signs, under these specific challenging conditions. Hence, to achieve their goal, two benchmark datasets (ITSRB and ITSDB) are proposed. These two datasets include more than 5800 images. For the obtained results, the proposed network (PFANet) reaches an accuracy of 93.57% and 97.21% onITSRB, and GTSRB [6] respectively, and that without data augmentation (using two NVIDIA-P40 GPUs).

Opting instead for data augmentation, the work of Fang, Cao and Li [7] aims to ameliorate and improve the recognition of blurred and dark images. Hence, the lightweight network MicronNet-BF is developed. The approach is based on the small CNN MicronNet, with an objective of reducing the parameters of the original network from 0.51 to 0.44 million parameters. Factorization and batch normalization are used, and the $5 * 5$ filter is replaced with two $3 * 3$ convolutional kernels to ensure an efficient and accurate classification. To evaluate the performances of the network, GTSRB is used. The obtained results shows that MicronNet-BF outperforms MicronNet by achieving 99.383%, against 98.9% obtained by MicronNet on GTSRB using a Linux Ubuntu20.4 and GPU (GTX1080ti of NVIDIA). For BTSRD [8], the achieved performances reach 82.122% and 80.388%, for MicronNet-BF and MicronNet respectively.

Adopting also a light architecture, Naim and Moumkine [9] use a very small number of parameters through a very thin DL Network. The adopted CNN (LiteNet) is designed to classify traffic signs using a bank of filters for a more efficient extraction of representative features. To test the performances, GTSRB is used. Compared to state-of-the-art methods, the accuracy obtained using this public dataset is 99.15%, and that using NVidia K80 GPUs. While the number of parameters is about 2.10 million parameters.

A small CNN is also used by Li, Li and Zeng [10]. The network is based on extracting low and high dimensional features to improve the performances, and that using convolutional pooling for two networks, a traditional and an improved CNN architecture. For the traditional network (Tra-Net), global average pooling is used in addition to Normal Normalization after the convolution. For the improved CNN (MyNet), TS-Module are proposed for features extraction. After the training, the results show that the improved CNN (MyNet) achieves better performances, while it reduces the parameters by 264.3 KB. Furthermore, the network reaches an accuracy of 97.4% using GTSRB. For BTSRD, the obtained accuracy is about 98.1% with an inference time of 705.10 ms per image. The experiments are realized using 32 GB RAM, i7-6700K quad-core and eight-thread CPU and NVIDIA-GTX1070Ti GPU.

From the literature review [1], we find that the majority of the researches concentrate rather on the obtained accuracy, while just few works deal with the challenge of the speed and the inference time, while it represents in fact one of the three key factors of an efficient recognition of traffic signs. These factors include the accuracy, the real-world adverse conditions, in addition to the latency required to ensure the real-time recognition. Furthermore, within the works that treat the inference time of DL approaches, hardware optimization is usually used to ameliorate the computation speed. However, that makes on the one hand the process of training, validation and testing very expensive, and

limits on the other hand the use of this type of approaches in the environments that are characterized by low and limited resources.

Another important issue that should be underlined is that the literature review shows also that many lightweight architectures perform very well in comparison to heavy and deeper ones. Hence, making these networks much faster is the challenge that should be faced by researchers, in order to ensure the implementation of this type of networks in real-time applications, and benefit by consequence from their high and outstanding performances. From this perspective, the next section presents the adopted approach proposed to find some elements of answers to these faced challenges.

3 Adopted Approach

In fact, the computational complexity of CNNs is related to many factors. One of these important factors is the type of the used layers, especially the hidden ones. In fact, we find that the hidden layers used in CNNs are generally fully-connected. However, this type of layers is very computationally expensive, due to the huge number of nodes, and operations needed to transform each feature map. To reduce this computational load, the convolutional layers are generally used as the first layers of the network before applying the full connections. These convolutional layers are essentially based on multidimensional treatment of data. On the other hand, we find that the 1D standard Convolutional networks gain recently more attention and popularity in many areas of research [11], especially in the medical field, as for example the classification of biomedical data, anomaly detection, early disease diagnosis, etc. [12]. In effect, the popularity of this networks is due to their low-cost implementation and computational complexity compared to the traditional 2D CNNs. Which makes them very suitable for real-time applications, and limited computational power devices, especially mobile ones.

Accordingly, and in order to benefit from the advantages of these two types of networks (1 and 2D CNNs), instead of merging the networks as proposed by Alkhatib, Hafiane and Vieyres [13], our first approach consists on using 2D convolutional layers for features maps' extraction, while applying 1D convolutions in the hidden layers instead of the fully-connected ones. For this purpose, we have used two networks, a standard CNN (Dense CNN) and a CNN with 1D convolutions in the hidden layers (Conv 2&1D Network). To train the two networks, we have used two extracted sets from the Challenging Unreal and Real Environments for Traffic Sign Recognition database (CURE-TSR) [14]. The first extracted set includes five classes of traffic signs (Fig. 1), with 969 images for the training, and 3769 for the testing set. For the used networks, they include both two convolutional layers with 32 and 64 kernels of $3 * 3$ and $5 * 5$ size. For the hidden layers, the first one includes two dense layers of 250 and 500 nodes respectively, while the second network includes instead one-dimensional convolutional layer with 10 kernels followed by a max pooling of 4. Hence, after 10 epochs of training the obtained results are presented in Table 1.



Fig. 1. Types of traffic signs in the 1st dataset.

Table 1. Obtained results with the 5 classes set.

Network type	Architecture	Parameters	Epoch	Inference	Accuracy
Dense CNN	Conv 32/64 and Dense 250/500	10,996,415	10	7 ms/step	97.26%
Conv 2&1D	Conv 32/64 and Conv1D 10/Max 4	592,945	10	5 ms/step	98.51%

For the second used database, it contains six additional classes with more challenging types of signs for a total of 11 classes (Fig. 2). The number of training set is 1331 images affected by a certain number of adverse conditions (rain, snow, blur, etc.). Table 2 shows the obtained results after the training using the 11 classes set for 10 epochs.



Fig. 2. CURE-TSR extracted training set (11 classes).

Table 2. Obtained results with the 11 classes set.

Network	Architecture	Parameters	Epoch	Inference	Accuracy
Dense CNN	Conv 32/64 and Dense 250/500	10,999,421	10	7 ms/step	95.27%
Conv 2&1D	Conv 32/64 and Conv1D 10/Max 4	1,241,851	10	6 ms/step	96.90%
	Conv 32/64 and Conv1D 10 and Slide 10/Max 4	171,191	10	2 ms/step	95.93%

The obtained results from Table 1 and Table 2 show that, applying one-dimensional convolutions in the hidden layers of standards CNNs instead of the fully-connected ones has a positive impact on increasing their efficiency (accuracy and speed). Hence, we find that the number of used parameters decrease considerably from 10 million to less than 1.3 million using the two datasets (Fig. 3). Furthermore, it improves also the inference time while increasing the obtained accuracy, and that by more than 1%. We find also

that, increasing the size of the moving window of the convolutional filters decreases enormously the number of used parameters to just 172k, while it speeds up the inference time to 2 ms by image. Which presents another advantage of 1D hidden convolutions. Accordingly, to further improve the obtained performances, we have adopted a second approach based on applying a 1D final module using instead a different type of 1D layers. Extracting more meaningful information from the final vector is the objective of the proposed approach, and that by adding a 1D module based essentially on creating 3 channels for this final vector, while applying different sets of kernels to each of the channels. This added module consists hence on a set of one-dimensional separable convolutions to get the final predictions.

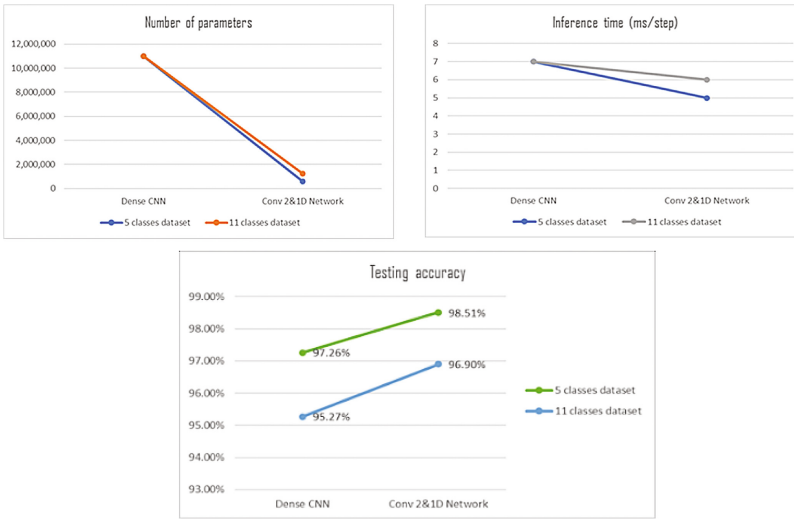


Fig. 3. 1D Conv impact on the obtained accuracy.

These vectors are then mixed and multiplied using 20 filters. In order to reduce the size of the output vectors, one-dimensional max pooling is applied. Finally, the vectors are transformed to a one big vector to be feed to the final prediction layer. The objective of concatenating these vectors, after the separable convolutions, is to extract more representative features from this final vector (Fig. 4). Which will help to enhance the network performances, while speeding the processing time at the same time. The next section shows the testing results of the adopted approach.

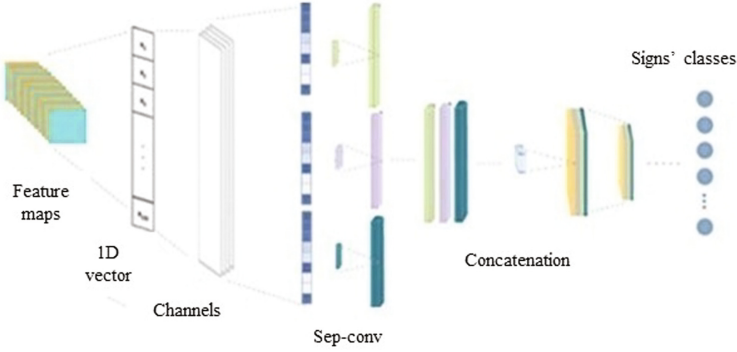


Fig. 4. One-dimensional final module.

4 Testing Results

To test the performances of the adopted approach, we have added the one-dimensional final module to our networks Map-CNN and Mean-LC [15, 16]. The experiments are conducted using 3.60 GHz, i3-8100 CPU, 4GB RAM.

4.1 Map-CNN's Testing Performances

For Map-CNN [15], it is a network that is based on a limited number of parameters and two-dimensional data in the hidden layers, and that using partial connections or two-dimensional dense layers. To train hence the network, we have used first the five classes dataset (Fig. 1). The accuracy and loss curves obtained during each epoch of the training process are presented in Fig. 5. For the testing stage, we have used the 3769 images of the five classes testing set. Hence, after the testing process, the obtained results show that applying the final one-dimensional module to Map-CNN has improved the obtained accuracy with almost 98.62% Table 3. To summarize the prediction results of the adopted approach, the confusion matrix of the testing set is presented in Fig. 6.

To evaluate the performances of the proposed approach to more challenging types of signs, we have used the extracted set with 11 classes. However, in order to accelerate the training, we have used instead transfer learning from the previous pre-trained five classes network to the 11 classes one. Hence, the general features learned by the 5 classes network will be reused for the 11 classes one, and that by freezing the first layers while unfreezing the last ones. Thing that will help the network to learn the specific features of the new signs when using this larger dataset.

Figure 7 and Fig. 8 show the performances of the training (90%) and the validation process (10%), using the 11 classes with transfer learning. The classification report of the precision, recall and F1-score, after the training process, is presented in Table 4. For the testing stage, we find that the obtained performances reach 97.67% in terms of accuracy, as presented in Table 5. After adding the final module to Map-CNN, and with the objective of further evaluate the performances of the proposed approach, we have added the one-dimensional final module to Mean-LC5, as presented in the next section.

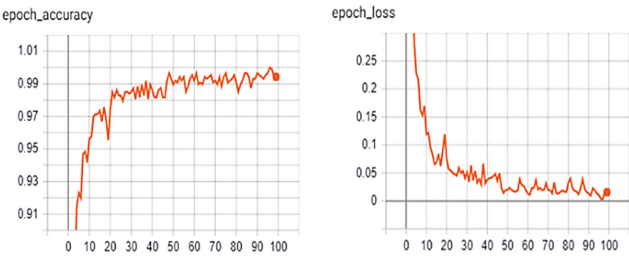


Fig. 5. Accuracy and loss curves with one-dimensional final module.

Table 3. Testing results.

Testing set	3769 images
Accuracy	98,62%

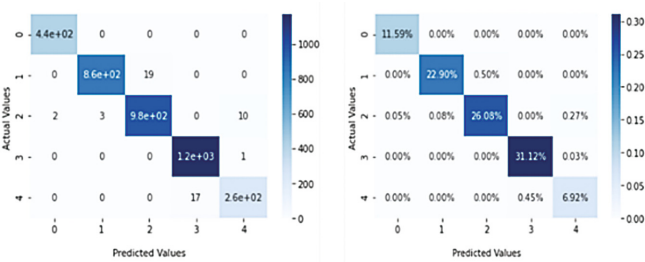


Fig. 6. Confusion matrix with labels and percentages.



Fig. 7. Accuracy curve during the training



Fig. 8. Loss curve during the training

Table 4. Classification report.

Classes	Precision	Recall	F1-score	Instances
0	1	1	1	145
1	0,99	1	0,99	171
2	1	1	1	232
3	0,99	1	1	147
4	1	0,99	1	176
5	1	0,99	0,99	87
6	1	1	1	61
7	1	1	1	61
8	0,98	1	0,99	40
9	1	1	1	37
10	1	0,95	0,97	40

Table 5. Testing results.

Testing set	4039 images
Accuracy	97,67%

4.2 Mean-LC5's Testing Performances

As for Map-CNNs, Mean-LC (Mean-LC 4&5) networks are based on keeping the multi-dimensionality of data in the hidden layers. In addition to that, the input layer of this type of networks is not a convolutional or a dense one, but it is instead a subsampling layer with separable convolutions. The objective of adding these layers is to reduce the size of input images without losing important details from the images. Unlike Map-CNN, we have added Dropout to randomly deactivate some nodes from the generated vectors before applying the one-dimensional convolutions in the final layers. Table 6 shows the number of the parameters used in the network after adding this final module. To compare the adopted approach with state-of-the-art methods, we have added the one-dimensional final module to Mean-LC5, and that using the public dataset BTSRD [8]. Hence, to train the network we have used BTSRD which includes 4575 images. We have trained the network with 4117 images (90%), and the other 10% are used for the validation process. After the training and the validation process, we have tested the performances of the adopted approach using the 2520 testing images of BTSRD Table 7.

Table 6. Number of parameters (Mean-LC5).

Total parameters	31 212
Trainable parameters	31 084
Non-trainable parameters	128

Table 7. Obtained Accuracy (testing set).

Testing set	2520 images
Accuracy	99%

According to the obtained results, we find that adding the one-dimensional final module has a positive impact on improving the obtained accuracies and accelerating the inference time. In comparison to the presented works from the related works’ section, we find that, the network proposed by Mehta, Paunwala, and Vaidya [3] reaches an accuracy of 97.06%, however it uses an important number of parameters that exceed 10 million, and that using BTSRD. For MicronNet-BF and MicronNet [7], the accuracy obtained is 82.122% and 80.388% respectively using the same dataset with GPU, while the number of used parameters is more than 0.4 million for both networks. The highest performance from the presented works is achieved by the improved CNN MyNet [10], with an accuracy of about 98.1% using BTSCD, while the inference time of the network is 705.10 ms per image using GPU. Hence for the inference time, we find that the adopted approach using Mean-LC5 network with the final module outperforms state-of-the-art methods, specifically those designed for real-time classification of traffic signs. Furthermore, it ensures also a fast training and validation process, using limited resources’ systems. Moreover, the proposed approach reaches high performances in terms of accuracy, and that using the challenging set of CURE-TSR, and BTSCD.

5 Conclusion and Perspectives

Basing the recognition process on the hardware optimization is no longer enough, especially for limited resources environments. Which leads to consider even more the algorithmic optimization, in order to speed up the process. From this perspective, the objective of the proposed approach is to reduce this complexity. Hence, to achieve this goal, we have replaced the final fully-connected layers with a one-dimensional final module, based instead on a very limited number of parameters. The proposed networks from the adopted approach are essentially based on two principal modules: two-dimensional module, followed by a one-dimensional one based on 1D separable convolutions and subsampling. To evaluate the proposed approach, we have used different challenging types of signs with Map-CNN and Mean-LC5 networks. Therefore, the obtained results show that adding the final module ameliorates the classification accuracy, and accelerate considerably the inference time. Furthermore, the processing time during the training process is very convenient, which makes the validation process even less complex and faster using limited resources environments. The latency of the proposed approach is also very low, and doesn’t exceed 0.3 ms per image. Which makes the proposed approach very fast, while it reaches a very high accuracy using BTSRD (99%).

References

1. Bousarhane, B., Bensiali, S., Bouzidi, D.: Road signs recognition: state-of-the-art and perspectives. *Int. J. Data Anal. Tech. Strateg. Spec. Issue Adv. Appl. Optim. Learn.* **13**(1/2), 128–150 (2021)

2. LeCun, Y., Bengio, Y., Hinton, G.E.: Deep learning. *Nature* **521**, 436–444 (2015)
3. Mehta, S., Paunwala, C., Vaidya, B.: CNN based traffic sign classification using adam optimizer. In: *International Conference on Intelligent Computing and Control Systems* (2019)
4. Suriya, P.A.: Traffic sign recognition using deep learning for autonomous driverless vehicles. In: *Fifth International Conference on Computing Methodologies and Communication* (2021)
5. Zhou, K., Zhan, Y., Fu, D.: Learning region-based attention network for traffic sign recognition. *Sensors* (21), 686 (2021)
6. Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C.: The german traffic sign recognition benchmark: a multi-class classification competition. In: *International Joint Conference on Neural Networks* (2011)
7. Fang, H., Cao, J., Li, Z.: A small network MicronNet-BF of traffic sign classification. *Comput. Intell. Neurosci.* (2022)
8. Mathias, M., Timofte, R., Benenson, R., Gool, L.: Traffic sign recognition — how far are we from the solution? In: *International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, USA (2013)
9. Naim, S., Moumkine, N.: LiteNet: a novel approach for traffic sign classification using a light architecture. In: Bennani, S., Lakhri, Y., Khaissidi, G., Mansouri, A., Khamlichi, Y. (eds.) *WITS 2020. LNEE*, vol. 745, pp. 37–47. Springer, Singapore (2022). https://doi.org/10.1007/978-981-33-6893-4_4
10. Li, W., Li, D., Zeng, S.: Traffic sign recognition with a small convolutional neural network. *IOP Conf. Ser. Mater. Sci. Eng.* (688) (2019)
11. Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.: 1D convolutional neural networks and applications: a survey. *Mech. Syst. Signal Process.* **151** (2021)
12. Wu, J.X., Pai, C.C., Kan, C.D., Chen, P.Y., Chen, W.L., Lin, C.H.: Chest X-ray image analysis with combining 2D and 1D convolutional neural network based classifier for rapid cardiomegaly screening. *IEEE Access* **10**, 47824–47836 (2022)
13. Alkhatib, M., Hafiane, A., Vieyres, P.: Merged 1D-2D deep convolutional neural networks for nerve detection in ultrasound images. In: *25th International Conference on Pattern Recognition* (2021)
14. Temel, D., Kwon, G., Prabhushankar, M., AlRegib, G.: CURE-TSR: challenging unreal and real environments for traffic sign recognition. In: *31st Conference on Neural Information Processing Systems (NIPS)*, Machine Learning for Intelligent Transportation Systems Workshop, Long Beach, CA, USA (2017)
15. Bousarhane, B., Bouzidi, D.: Map-CNNs: thin deep learning models for accelerating traffic signs recognition. *Adv. Dyn. Syst. Appl. (ADSA)* **16**(2), 1777–1798 (2021)
16. Bousarhane, B., Bouzidi, D.: New deep learning architecture for improving the accuracy and the inference time of traffic signs classification in intelligent vehicles. In: Lazaar, M., Duvallet, C., Touhafi, A., Al Achhab, M. (eds.) *BDIoT 2021. LNNS*, vol. 489. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-07969-6_2
17. Bousarhane, B., Bouzidi, D.: Partially connected neural networks for an efficient classification of traffic signs. In: *30th IEEE Conference of Open Innovations Association FRUCT*, University of Oulu, Finland (2021)
18. Tan, M., Le, Q.: EfficientNet: rethinking model scaling for convolutional neural networks. In: *36th International Conference on Machine Learning*, Long Beach, California (2019)
19. Xu, S., Niu, D., Tao, B., Li, G.: Convolutional neural network based traffic sign recognition system. In: *5th International Conference on Systems and Informatics* (2018)
20. Larsson, F., Felsberg, M.: Using fourier descriptors and spatial models for traffic sign recognition. In: Heyden, A., Kahl, F. (eds.) *SCIA 2011. LNCS*, vol. 6688. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21227-7_23
21. Møgelmoose, A., Liu, D., Trivedi, M.: Detection of U.S. traffic signs. *Trans. Intell. Transp. Syst.* **16**(6), 3116–3125 (2015)



Deep Learning Approach for a Dynamic Swipe Gestures Based Continuous Authentication

Zakaria Naji^(✉) and Driss Bouzidi

Smart Systems Laboratory (SSL), National School of Computer Science and Systems Analysis (ENSIAS), Mohammed V University, Rabat, Morocco
`zakaria_naji@um5.ac.ma`, `driss.bouzidi@ensias.um5.ac.ma`

Abstract. The amount of sensitive data stored on mobile devices has increased. Current mobile device security schemas, such as pins, passwords, patterns, or even physiological biometrics are not secure enough to protect these data. There are several continuous or active authentication approaches that would provide an additional line of defense, designed as a security countermeasure. This paper introduces a Continuous Authentication system based on the swipe gestures as images. Therefore, we designed a Dual Input Model based on MobileNetV2. For training and testing the model, we used the two public Datasets, BioIdent and HMOG. As a result, our model achieved an EER of 10.45% which represents a good rate compared to the results of existing research.

Keywords: Continuous Authentication · Behavioral biometrics · Swipe gesture · Mobile device · Deep Learning

1 Introduction

These days, smartphones are becoming important tools in our daily life. As far, there are millions of people around the world who use connected smartphones. These smartphones have evolved from simple mobile devices to mini-computers with complex systems that connect to large numbers of communication networks and provide many services such as voice calling, messaging, mailing, video conferencing, multimedia sharing, online banking, and data synchronization [1]. These features and services store sensitive information on the device. Therefore, how to authenticate and validate the identity of smartphone users when they use them.

Static user authentication methods on mobile devices define an entry point into the system. Typically, the users face a password challenge and only get access if they enter the correct password. Static Authentication in mobile phones is divided into two types: The first type is Knowledge-Based, it uses common knowledge between the user and the system as identity-related information for authentication. The knowledge-based secret can be in the form of text such as PINs and alphanumeric passwords or graphics like gesture patterns. The other type is biometrics-based. Compared to knowledge-based authentication, authentication based on physiological biometrics is more convenient (i.e. no need to

memorize secret code) and more secure (i.e., it is more difficult to steal), as it exploits the unique human biometric characteristics like fingerprints and face features that are inherent in parts of the user body for authentication [2].

Although these methods may be convenient for the user, they may also be susceptible to many attacks to bypass the static authentication methods. For example, PINs and Passwords are susceptible to attacks like shoulder surfing, reflections [3], or side-channel attacks [4]. Patterns are also susceptible to attacks such as smudge attacks [5]. Even biometrics are bypassed using manipulation-based attacks like biometric template attacks [6]. Added to this, by leaving the mobile device unattended and unlocked for longer or shorter periods, anyone can have access to the same resources as the genuine user.

To overcome these challenges, Continuous Authentication (CA) must be applied to respond imperceptibly to constantly monitoring the identity of the user by comparing current usage data to that of the genuine user. CA is based on the behaviors of the user during his interaction with the device. Behaviors are specific and unique when using a device to identify users from each other and ensure the right person is behind the screen. This is by integrating the most discriminating aspects of his behavior, such as his way of typing, tapping, swiping, or moving the phone. Using behavioral biometrics as an authentication method can be secure. Instead of having to memorize a password to unlock a mobile device, a behavioral biometric system can authenticate a user by simply interacting with the device, which can be much more convenient.

Therefore, in this paper, we propose a CA system that is based on behavioral biometrics by focusing on swipe gestures. These swipe gestures will be transformed into images and then processed by Deep Learning (DL) technique specifically Convolutional Neural Network (CNN) in order to distinguish the swipe gestures of genuine users and swipe gestures of imposter users. The reason for using the swipe gesture as an image is for taking advantage of the strength of CNN and its ability to detect distinct and significant features from images with a high degree of accuracy. However, deep learning models require higher computing resources to minimize the time consumed while training or inferencing [7,8]. In the remainder of the paper, Sect. 2 discusses related work. Section 3 outlines the proposed system. The evaluation results are presented in Sect. 4. In Sect. 5, we conclude our study.

2 Related Works

Continuous Authentication (CA) can complement the static authentication methods by monitoring the user's interactions after a successful login. Authentication based on behavioral biometrics has attracted a lot of attention in recent years. It leverages behavioral biometrics that captures the user's unique characteristics or habits for authentication. For example, when the user taps or swipes on the touch screen, finger movements exhibit a unique behavior pattern that can be used for authentication. Similarly, walking and talking also result in unique patterns that can be used to distinguish users. Generally, authentication

based on behavioral biometrics is more acceptable to users than authentication based on physiological biometrics. Indeed, it represents less private information compared to unchanged body features, and users are less cautious in using them.

The authors in [9] introduced three modalities for continuously authenticating smartphone users: swiping gestures, typing behavior, and phone movement patterns. The authors developed a mobile application that consists mainly of a browser named Lightning Browser and a service that runs in the background to collect data. The application allowed authors to extract segments of phone movement patterns (i.e. accelerometer readings) corresponding to individual swipe gestures. In addition to the data extracted from the swipe gestures, they extracted an additional three seconds of phone movement patterns that were generated before and after each swipe gesture. To develop a CA system for smartphone users, the authors propose a multimodal framework that uses more than one modality (depending on their availability) and combines them to cover the entire interaction of the user with the phone. The fusion of modalities is pursued by two approaches, Feature Level Fusion (FLF), or Score Level Fusion (SLF). For classification, the authors used k-NN ($k=11$) with an Euclidean distance and Random Forest with a thousand trees. As result, the Random Forest classifier achieved the highest accuracy rate, 93.33% when using FLF of swipe gestures and phone movements while swiping.

In [10], The authors introduce the HMOG (Hand Movement, Orientation, and Grasp) dataset. It contains data from both the touch-screen and sensors. 100 volunteers are surveyed over 24 sessions to gather data from touchscreen and collecting features like x-y coordinates, finger-covered area, pressure, and so on. In addition to touchscreen data, the authors propose a new set of features derived from micro-movements and obtained from accelerometer, gyroscope, and magnetometer sensor. This data is generated when users interact with the touchscreen. As a result, they achieved an EER of 15.1% by combining HMOG and tap features.

Another research [11] presented a CA system based on a multi-modal that makes use of multi-modal behavioral biometrics, specifically touch dynamics and phone movement. The authors combined two publicly available datasets, HMOG [10] and BioIdent [12], which were collected using touchscreen, accelerometer, gyroscope, and magnetometer sensors. For classification, they used a variety of machine learning algorithms, including Random Forest, Support Vector Machine, and K-Nearest Neighbors. As a result, the Random Forest classifier achieved an EER of 13.56%.

3 Our Continuous Authentication System

Our CA system is based on using the swipe gesture as an image and creating a model that is able to distinguish between the swipe gesture of the genuine user and the swipe gesture of the impostor user. The system architecture is composed of four main phases as shown in Fig. 1. It includes Data collection and Preprocessing, Data Splitting, Training, and Authentication phases. This section describes the approach we used to conduct the proposed system.

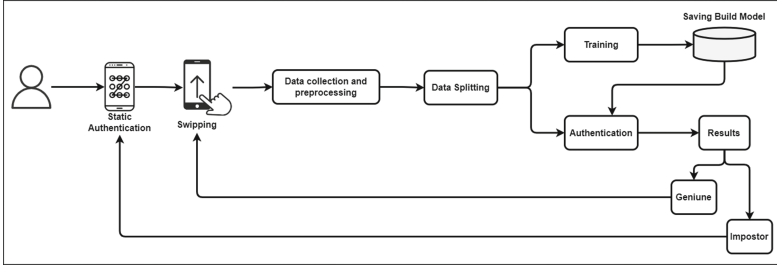


Fig. 1. Our proposed system architecture

3.1 Datasets Description

In order to compare with other studies, we tend to use a dataset that contains touch events data and then transform the swipe gestures into images. Thus, we choosed to use two publicly available datasets: the HMOG dataset [10], and the BioIdent dataset [12].

The HMOG dataset contains data from both the touchscreen and other sensors including the accelerometer, gyroscope, and magnetometer. The dataset was collected from 100 users by performing three different tasks using a Samsung Galaxy S4 smartphone (reading a document, entering text, and using a map application). The collection experiments were based on four scenarios: reading a document while sitting, reading while walking, browsing while sitting, and browsing while walking. Since our study focuses on the swipe gesture, we used only data collected from touch events. The features extracted from the touch events are the following elements: Systime, EventTime, ActivityID, Pointer_count, PointerID, ActionID, X, Y, Pressure, Contact_size, Phone-orientation.

The BioIdent dataset was created based on a set of touch events data collected from 71 users by performing two different tasks and using 8 mobile devices of different resolutions. the vertical swipe gestures were created by reading a text, whereas the horizontal swipe gestures were created by swiping an image gallery. The raw touch data generated a list of data with the following elements: user_id, doc_type, action, phone_orientaton, x_coor, y_coor, pressure, finger_area.

In this study, we combined two subsets from HMOG and BioIdent. As result, each user in the combined dataset has 100 vertical swipe gestures and 100 horizontal swipe gestures.

3.2 Dataset Preprocessing

The data of swipe gestures was collected as touch events data. In our approach, we used the swipe gesture as an image. So, in this section, we will represent the process of preprocessing the data as shown in Fig. 2.

In almost all studies, the features extracted from the swipe gestures are mostly represented by: the first and the last point of the swipe, the average pressure while swiping, the length of the swipe gesture, and the velocity of swiping. In our approach, we proceed toward transforming these swipe gestures into

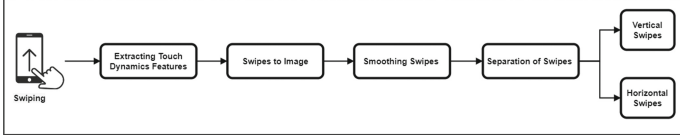


Fig. 2. Dataset preprocessing

images. A swipe gesture is made up of several points. Each point has an x coordinate, a y coordinate, and an action type. The action type has three values: 0 (represents the first contact with the touchscreen), 1 (represents the last contact with the touchscreen), and 2 (represents the movement of the finger on the touchscreen while swiping). So to create a swipe gesture we connected the points to each other in the following order: the swipe gesture starts with a point that has an action 0 and then followed by several points that have action 2 and end up with a point that has an action 1 (see Fig. 3).

After creating the swipe gesture as an image, we found that it is not smooth enough. Therefore, to smooth the swipe gesture as shown in Fig. 3, we used the cubic B-Spline interpolation curves by using the function (`scipy.interpolate.splrep()`) in Python programming [13]. The advantage of the cubic-B spline interpolation curve is that it can accommodate various shapes with only four parameters, in addition, the curve changes smoothly [14].

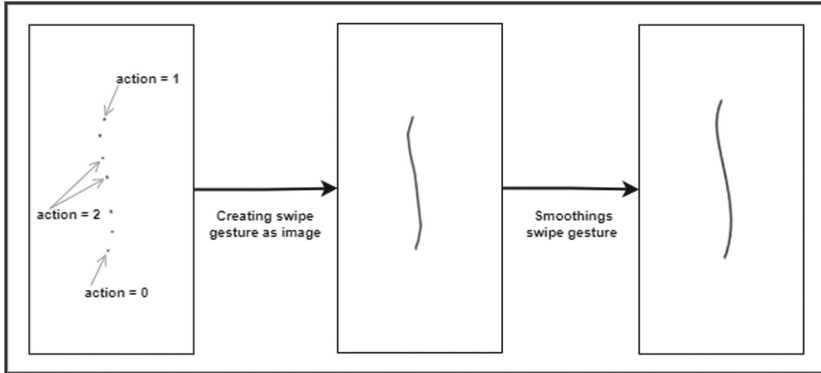


Fig. 3. Creating and Smoothing Swipe gesture

We divided Swipe gestures into two types, horizontal swipe gesture (Hs) and vertical swipe gesture (Vs). The separation is based on the angle θ between the line $[(x_0, y_0), (x_n, y_n)]$ and the line $[(x_0, y_0), (x_n, y_0)]$ of a swipe gesture as shown in Fig. 4. If θ is less than 25° , the swipe gesture is considered Hs, and if θ is greater than 25° the swipe gesture is considered Vs. To calculate the angle θ between two lines, we used the Eq. (1).

$$\theta = \arctan \left(\left| \frac{M_2 - M_1}{1 + M_1 \times M_2} \right| \right) \quad (1)$$

M_1 and M_2 are the slope value of the two lines.

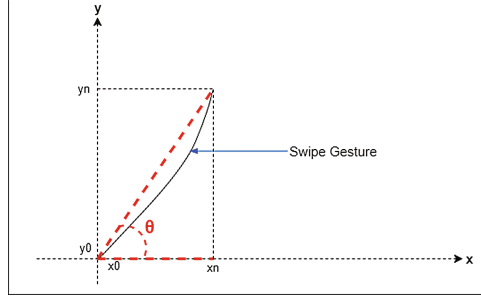


Fig. 4. Calculating angle of swipe gesture

3.3 Data Splitting

The combined dataset is composed of data from 51 users. Each user has 100 samples. A sample is composed of a vertical and a horizontal swipe gesture. For each user's dataset, we have two classes, a swipe gesture that belongs to the genuine user and another one that belongs to an impostor user.

A user training set includes 80 samples from his samples that serve as genuine's and an equal number of samples drawn from the remaining 50 users and serve as impostor's samples. To avoid bias in favor of either the genuine or impostor classes, the amount of samples is almost equal for both the genuine and impostor classes. For testing, we used the user's remaining 20 samples that serve as genuine's and one sample from each of the 50 remaining users that serve as impostor's samples.

3.4 Training

The intended goal of the proposed CA system is to continuously authenticate a user based on his swipe gesture behavior patterns. The adopted model (see Fig. 5) will be trained over the training set in order to be able to distinguish between the swipe gestures of the genuine user and the swipe gestures of the impostor users. The model architecture will be constructed using MobileNetV2 as a feature extractor.

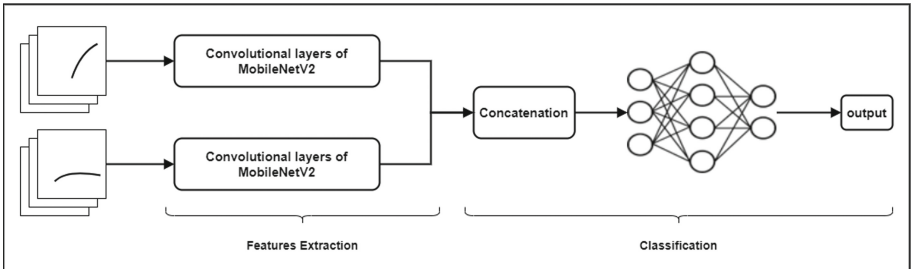


Fig. 5. Dual input model

Recently, many improvements have been made in exploring the architecture of several CNN architectures in order to reduce the number of hyperparameters without affecting the performance. A substantial amount of work has also gone into modifying the connectivity structure of internal convolutional blocks like in ShuffleNet [15] and other CNNs.

The design of our DL network is based on MobileNetV2 which is an efficient and lightweight neural network architecture. The basis of the network is the Depthwise Separable Convolution layer which reduce the computation and memory requirements when compared to traditional convolutional layers. According to [16], the parameters used have been reduced from 4.24 million to 3.47 million, but with an improved accuracy. Furthermore, MobileNetV2 utilizes a method known as linear bottlenecks, which enhance the efficiency of the model by minimizing the number of channel-wise computations. As a result, MobileNetV2 is frequently used for tasks such as image classification on devices with limited resources.

Our DL model called Dual Input Model is composed of two inputs (see Fig. 5). Each input has its own feature extractor which is MobileNetV2. The first Input is used for extracting features from vertical swipe gestures and the other one is used for extracting features from horizontal swipe gestures. The output of both inputs are concatenated and fed to a Flatten layer in order to transform the output of the concatenation layer into a one-dimensional feature vector. The Flatten layer is then followed by three fully-connected layers. They are composed of 16, 32, and 64 nodes respectively, with ReLU (Rectified Linear Unit) as an activation function. The output layer is made up of 2 with SoftMax as an activation function.

The base model MobileNetV2 uses the pre-trained weights, and its layers are partially finetuned. Therefore, in the training process, only the weights of the last layers of the base model and the head model were trained. To meet the input requirement of the MobileNetV2 model, all swipe gesture images are resized to 224×224 pixels. The optimization algorithm used in the training process is Adagrad, the reason to use it is that it outperforms the other DL optimization algorithms in a variety of disciplines such as vision classification [17]. The batch size is 20. We used a learning rate equal to 0.0001. To reduce the loss during the training process, we used categorical cross-entropy as a cost function.

4 Evaluation Results

4.1 Evaluation Metrics

For the evaluation of the model, we used the following metrics: accuracy, precision, recall, F1-Score, FAR, FRR, and Equal Error Rate (EER). Accuracy is the percentage of correct predictions made by our model. Precision refers to the number of true positives divided by the total number of positive predictions (i.e., the number of true positives plus the number of false positives). The recall or True Positive Rate (TPR) is calculated as the ratio of positive samples that were correctly classified as positive to the total number of positive samples, it measures its ability

to detect positive samples. The more positive samples detected, the higher it is. The F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. F1 is usually more useful than accuracy, especially in uneven class distribution. The False Acceptance Rate (FAR) is the percentage of impostor users that the system incorrectly considered as genuine users. And the False Rejection Rate (FRR) is the percentage of genuine users that the system incorrectly considered as impostor users. Error Equal Rate (EER) is the common value when FAR and FRR are equal. We use EER because it accounts for the trade-off between FAR and FRR.

4.2 Results

According to Table 1, for user 29, the model achieved an EER of 0% as the best result. The model was able to distinguish between the swipe gestures of user 29 and those of the other impostor users. On the other hand, for user 19, the model achieved an EER of 38.46% and that is because the model incorrectly considered 60% of swipe gestures of the impostor users as they are belonging to the genuine user, even if it was able to only consider 4% of swipe gestures of the genuine user as swipe gestures of the impostor users. Overall, the model achieved an average EER of 10.45% for all users.

Table 1. Results

	Accuracy	Precision	Recall	F1 score	FAR	FRR	EER
User 29	100%	100%	100%	100%	0%	0%	0%
User 19	80%	80%	40%	53.33%	4%	60%	38.46%
Mean results	92.24%	84.71%	90.2%	86.93%	6.94%	9.8%	10.45%

The model showed encouraging results as shown in Fig. 6 which represents a summary of EER for all users. The figure depicts the median EER (center green line), the 25%, and the 75% percentiles. Their values respectively are 9.26%, 5.94% and 12.33%. Outlying users are individually reported outside the whiskers as black circles.

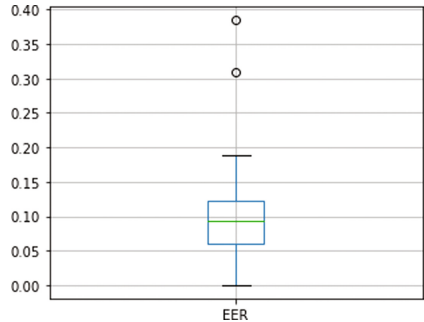


Fig. 6. Summary of EER for all users.

Our model was able to reach an EER of 10.45% which is an improvement in comparison with other works like [9–11], and [18] as shown in Fig. 7. Although, we were able to improve the performance of the model because of the use of the power of the CNN, the training time was very long. To the best of our knowledge, this is the first implementation and testing with swipe gesture as an image using CNN.

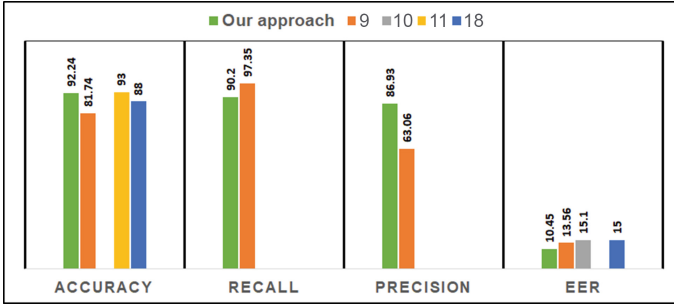


Fig. 7. Comparison of our evaluation results with related works.

5 Conclusion and Future Work

Mobile device user authentication is a critical technology that prevents unauthorized access to a mobile device or a mobile application to protect sensitive user information. In order to provide this protection service continuously, we proposed a CA system based on swipe gestures as images. Our system was divided into four primary stages. The first is Data Collection and Preprocessing, followed by Data Splitting, then the Training and at the end the Authentication phase. As a result, our proposed system achieved an EER of 10.45%, which is a promising result.

For future work, we will improve our proposed system through other functionalities such as the use of decentralized learning which is a technique that trains a model across multiple decentralized edge devices or servers holding local data samples without sharing them in order to provide privacy for the users.

References

1. Majeed, K., et al.: Behaviour based anomaly detection system for smartphones using machine learning algorithm. Ph.D. thesis, London Metropolitan University (2015)
2. Wang, C., Wang, Y., Chen, Y., Liu, H., Liu, J.: User authentication on mobile devices: approaches, threats and trends. *Comput. Netw.* **170**, 107118 (2020)

3. Xu, Y., Heinly, J., White, A.M., Monroe, F., Frahm, J.M.: Seeing double: reconstructing obscured typed input from repeated compromising reflections. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, pp. 1063–1074 (2013)
4. Shukla, D., Kumar, R., Serwadda, A., Phoha, V.V.: Beware, your hands reveal your secrets! In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, pp. 904–917 (2014)
5. Aviv, A.J., Gibson, K., Mossop, E., Blaze, M., Smith, J.M.: Smudge attacks on smartphone touch screens. In: 4th USENIX Workshop on Offensive Technologies (WOOT 10) (2010)
6. Xi, K., Ahmad, T., Han, F., Hu, J.: A fingerprint based bio-cryptographic security protocol designed for client/server authentication in mobile computing environment. *Secur. Commun. Netw.* **4**(5), 487–499 (2011)
7. Li, J., Zhang, C., Cao, Q., Qi, C., Huang, J., Xie, C.: An experimental study on deep learning based on different hardware configurations. In: 2017 International Conference on Networking, Architecture, and Storage (NAS), pp. 1–6. IEEE (2017)
8. Bousarhane, B., Bensiali, S., Bouzidi, D.: Road signs recognition: state-of-the-art and perspectives. *Int. J. Data Anal. Tech. Strat.* **13**(1–2), 128–150 (2021)
9. Kumar, R., Phoha, V.V., Serwadda, A.: Continuous authentication of smartphone users by fusing typing, swiping, and phone movement patterns. In: 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–8. IEEE (2016)
10. Sitová, Z., Šeděnka, J., Yang, Q., Peng, G., Zhou, G., Gasti, P., Balagani, K.S.: HMOG: new behavioral biometric features for continuous authentication of smartphone users. *IEEE Trans. Inf. Forensics Secur.* **11**(5), 877–892 (2015)
11. Mallet, J., Pryor, L., Dave, R., Seliya, N., Vanamala, M., Sowell-Boone, E.: Hold on and swipe: a touch-movement based continuous authentication schema based on machine learning. In: 2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML), pp. 442–447. IEEE (2022)
12. Antal, M., Bokor, Z., Szabó, L.Z.: Information revealed from scrolling interactions on mobile devices. *Pattern Recogn. Lett.* **56**, 7–13 (2015)
13. Oliphant, T.E.: Python for scientific computing. *Comput. Sci. Eng.* **9**(3), 10–20 (2007)
14. Caglar, H., Caglar, N., Elfaituri, K.: B-spline interpolation compared with finite difference, finite element and finite volume methods which applied to two-point boundary value problems. *Appl. Math. Comput.* **175**(1), 72–79 (2006)
15. Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: an extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6848–6856 (2018)
16. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520 (2018)
17. Hassan, E., Shams, M.Y., Hikal, N.A., Elmougy, S.: The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study. *Multimedia Tools Appl.*, 1–43 (2022)
18. Volaka, H.C., Alptekin, G., Basar, O.E., Isbilen, M., Incel, O.D.: Towards continuous authentication on mobile phones using deep learning models. *Procedia Comput. Sci.* **155**, 177–184 (2019)



Skin Cancer Detection Based on Deep Learning Methods

Sara Shaaban¹, Hanan Atya¹, Heba Mohammed¹, Ahmed Sameh¹,
Kareem Raafat¹ (✉), and Ahmed Magdy²

¹ Egyptian Chinese College for Applied Technology, Ismailia 41511, Egypt
ugs.20073china@suez.edu.eg

² Electrical Engineering Department, Suez Canal University, Ismailia 41511, Egypt

Abstract. One of the most common and dangerous types of tumors/cancer is skin cancer. It is an abnormal growth of the skin. Skin cancer that is not detected early spreads to other organs. In the era of deep learning technology and computer vision, determining whether or not a patient has cancer has become more accessible because the earlier the diagnosis is, the more it reduces the risk or prevents the patient's death. Artificial intelligence has a lot of applications in healthcare, especially in skin cancer detection. In this paper, A convolutional neural network is used to make a diagnosis system more accurate in diagnosing whether the patient has cancer or not. HAM10000 is a Dataset which used to train the model, and it is used to classify between cancer and no cancer. MATLAB is a software tool we used when we train deep learning algorithms were used to make our diagnosis systems. To achieve the highest level of accuracy, two algorithms were used to optimize the parameters and apply them in GoogleNet, MobileNet-v2, NasNet mobile, SqueezeNet, Darknet19, VGG16, Xception, ShuffleNet, Inception-v3, resnet18, resnet50 and nasnetlarge. Then the result of each of them is recorded, and the best algorithm is selected. Xception is used to make a diagnosis system with 96.66% accuracy to classify between cancer and no cancer.

Keywords: Artificial intelligence · Skin cancer · CNN · HAM1000

1 Introduction

Cancer has sparked a lot of interest in recent years due to the damage and spread it has caused. Cancer is considered one of the leading causes of medical care, as it causes death and has many kinds, including lung, liver, and stomach cancer. Skin cancer is the worst kind, but it can be cured successfully where it is discovered early. Skin cancer is the most common cancer type that affects humans [1]. Generally, the early discovery of cancer is the most effective treatment for saving humans and decreasing medical and financial burdens. Cancer takes an interest in the search line for cancer [2]. Many software developers were interested in expanding the discovery in skin cancer detection applications [3]. In addition to decreasing workload and helping detect skin cancer [4], AI also helps detect cancer. AI is used in machines and programs. AI is considered a branch of computer science. This paper aims to provide an application for early cancer

detection based on AI in diagnosing and managing skin cancer by training the AI model with the dataset HAM10000 and testing it on some other pictures to see its accuracy if it is capable of diagnosis detecting early skin cancer [5]. Artificial intelligence (AI) is the ability of a computer to make decisions and perform tasks as if they were a human brain [6, 7], and a self-driving car is an example of AI. The computer system must account for all external data and compute it to drive in the right direction to avoid accidents. Also, artificial intelligence (AI) is a fast-evolving field, with research in this area moving at a breakneck pace [8]. All diagnosis systems in clinics and hospitals have been based on computer Vision [9]. Deep learning is a branch of AI that consists of a neural network with three or more layers. The neural network tries to simulate the behavior of the human brain by learning from vast amounts of data to make predictions and optimizations. Skin cancer detection is one application that can benefit from deep learning. Any patient can use this technique to diagnose whether their skin is injured or healthy without human intervention [10–12]. Deep learning includes models and algorithms, and convolutional neural networks (CNNs) are one of these algorithms [13]. CNN is an artificial neural network (ANN) and has more than three layers used in image recognition and classification. CNN can identify faces, individuals, street signs, diseases, and many other aspects of visual data [14, 15]. Many applications depend on deep learning in health care, such as diagnosis systems [16].

In this paper, CNN will be used to make diagnosis systems used in skin cancer detection [17]. CNN can learn from datasets used to predict and classify whether or not a person has cancer [18, 19]. The diagnosis system must be highly accurate because any error rate can worsen the patient's case or death. MATLAB uses two algorithms to optimize the accuracy by adjusting the values of the algorithms' parameters such as pixel range, number of layers, scale range, mini patch size, and learning rate to obtain the best result in accuracy. Applying them to GoogleNet, MobileNet-v2, NasNet mobile, SqueezeNet, Darknet19, VGG16, Xception, ShuffleNet, Inception-v3, resnet18, resnet50 and nasnetlarge. And calculating the accuracy result for each of them [20]. The accuracy of classifying images into the cancer or no cancer category by Xception is 96.66%, MobileNet-v2 is 95.5%, GoogleNet is 94.96%, SqueezeNet is 92.91%, Darknet19 is 90.16%, VGG16 is 90.66%, Inception-v3 is 95.76%, resnet18 is 93.8%, resnet50 is 95.66%, nasnetlarge is 92.9% and Shufflenet is 91.61% [21]. Xception is selected to detect whether the patient has cancer or no cancer in our diagnosis system. One of convolutional neural network technique which has 71 layers is Xception technique [22]. Using Xception, we will present a diagnosis system with 96.66% accuracy to classify between cancer and no cancer.

2 Proposed Methodology

The proposed methodology of Skin Cancer Detection using Xception. The proposed method is shown in Fig. 1 using a block diagram. Each block is clarified in detail, such as the dataset. The proposed system uses the HAM10000, then Segmentation dataset, and it has been segmented to cancer and no cancer. The dataset is segmented by using python code. Pre-processing position of images process must be non-uniform in several terms. Xception is used for classification. Since the dataset consists of about 10.000 images

that involve complexity, 80% of shots are considered by training the following 20% of images for testing. By the way, training has to happen to him through pre- processing and then concludes a model, and testing has to happen to him through a process of pre-processing and then using that model by making a condition that shows the training is finishing or not.

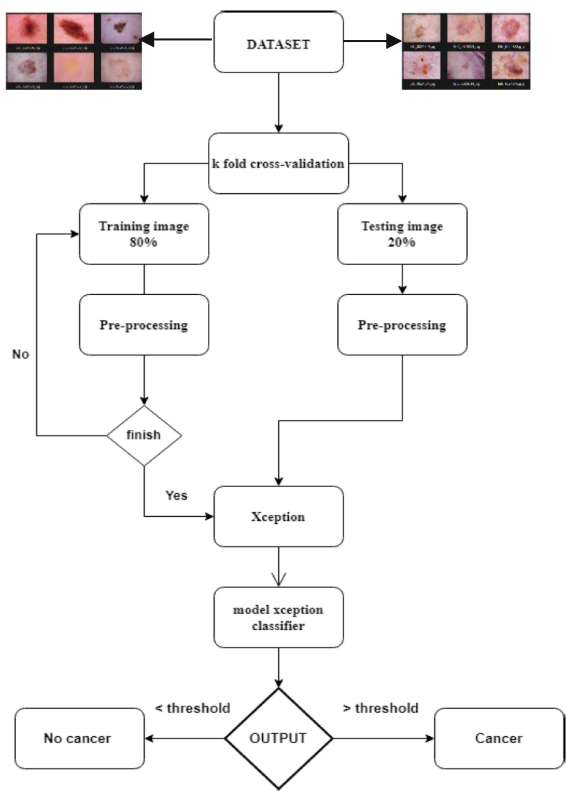


Fig. 1. System flow diagram.

3 Proposed Pretrained DNN Techniques

GoogleNet is a convolutional neural network consisting of 22 layers deep, including five inception layers as shown in Fig. 2, GoogleNet started in 2014 named “GoogleNet” as the team developed it at Google. Its goal was to achieve a new state of classification and detection in The ImageNet Large Scale Visual Recognition Challenge (ILSVRC), a large-scale evaluation of object identification and image categorization methods. Its role is to make researchers make comparison progress in detection and classification. And scored 93.33% accuracy in (ILSVRC). Classify images into 365 categories, such as field, park, runway, road, and lobby. These networks have learned to represent other

features for a wide range of images. One module comprises four 1×1 convolutions, one 3×3 convolution, one 5×5 convolution, and one 3×3 max pooling. We were using the GoogleNet algorithm in our paper to classify our data into types of cancer and no cancer.

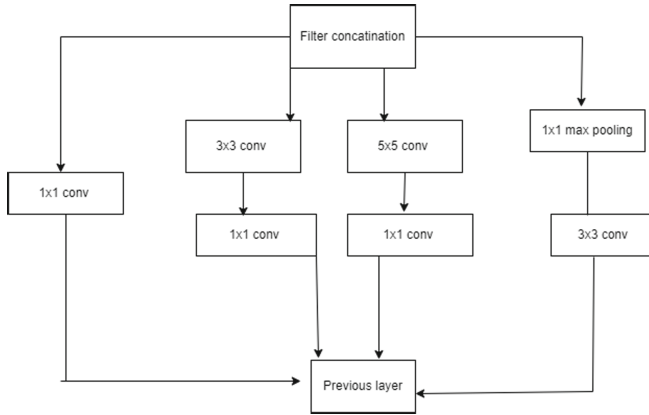


Fig. 2. GoogleNet algorithm.

Google design Mobilenet v1 [23] was designed by Google in 2015. The MobileNet v2 model was proposed by Google in 2017. It has image input size 224 by 224 model. The Mobilenet model was used to identify and classify the pests of *Lycium barbarum* [24]. This makes high progress in detecting planets and classifying many problems with planets. It's classified between 30 kinds of pests and more than ten diseases. MobileNet v2 is an upgraded model of Inception. It's faster than about three times this model score of 98.23%, according to research [25]. The model method is shown in Fig. 3; in our study, we got 95.5% in the classification of data to cancer and no cancer.

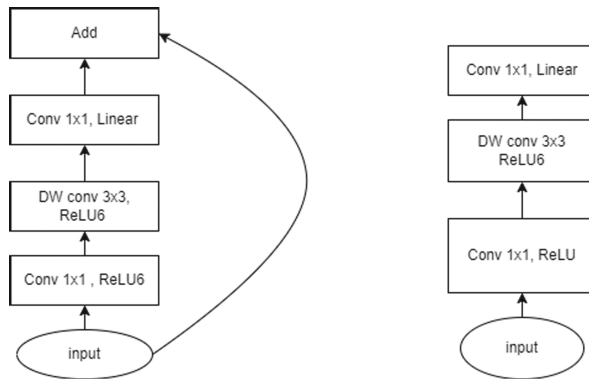


Fig. 3. Mobilenet v2 Convolution neural network.

The Nasnetmobile model is a convolutional neural network created in 2015 that is trained on millions of images; it can also classify over 1000 different objects into different categories, such as bottles or animals. It has a size of 224, consisting of about 4 million parameters. Nasnetmobile has fewer parameters than other models and has been used in the field of cancer detection to distinguish between 7 types of skin cancer and achieve 97.90% accuracy [26]. It achieves 93.110% accuracy in this paper to distinguish between cancer and no cancer cases.

SqueezeNet is a model of convolutional neural network consisting of 18 layers released in 2016. It can be trained on a million images and classify objects into 1000 object categories. Each input size is 227 by 227. This model scores high accuracy in image classification and detection, especially used in fingerprint detection to distinguish between a live fingerprint and a fake one [27]. As shown in Fig. 4, it shows how SqueezeNet works and how pre-trained models work. In our research, SqueezeNet scored 92.91% accuracy in classifying it as cancer or no cancer.

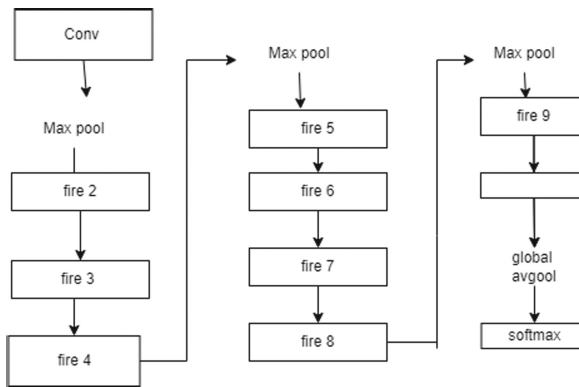


Fig. 4. SqueezeNet architecture.

Darknet19, model of convolutional neural network consisting of 19 layers trained on millions of images from ImageNet can classify images into 1000 objects. Darknet19, used in image classification of tobacco leaves, scored a high accuracy of 97% [28]. It's consisted of 5 max pooling and 1 softmax, input size 256 by 256, and 64 input layers. It's totally consisted of three processes: extraction features, classification layer, and clustering.

ShuffleNet, a convolutional neural network model trained on more than a million images from the ImageNet dataset, was released in 2018. It consists of a 3×3 convolutional layer, Channel shuffle, global average pooling, FC layer, and Leaky ReLU, as shown in Fig. 5 [29]. Shufflenet builds convolutional blocks by using more features in the input channels. This project has achieved a 91.61% accuracy in classifying cancer and no cancer.

Inception v3 is a convolutional neural network consisting of 48 layers deep trained on millions of images from ImageNet. It can classify images into 1000 objects. It has an image input size of 299 by 299 pixels. Inception v3 is based on three convolution layer

inception modules and classifiers. The first version of this consisted of 1×1 convolutions, 3×3 convolutions, 5×5 convolutions, as shown in Fig. 6. The model developed and has a better performance, gives higher efficiency, and has a deeper network. It makes steps to improve the efficiency by factorization to smaller convolutions, spatial factorization onto asymmetric convolutions, Utility of Auxiliary classifier, efficient Grid Size Reduction It has the lowest error rate of 3.6 percent among the compared models [30]. And they scored a high accuracy in breast cancer detection at 99.928% [31].

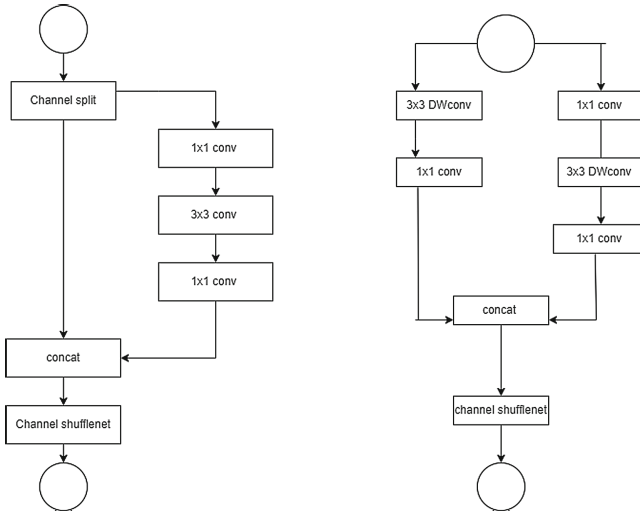


Fig. 5. Shufflenet convolution neural network.

Xception is a deep convolutional neural network consisting of 71 layers deep. The model network is trained on more than a million images from ImageNet. The network can classify more than 1000 objects. It has an image input size of 229 by 299 pixels as shown in Fig. 7. Xception is based on inception It stands for extreme inception; it's a working method. The Xception score has a high accuracy in detection of healthy and not healthy planets, reaching 97.5% and trained on about 70000 images [32].

VGG-16 is a convolutional neural network consisting of 16 layers deep trained on millions of images from ImageNet. It can classify images into 1000 objects. It has an image input size of 224 by 224. This model achieves a high accuracy in image classification with 92.7% on ImageNet. VGG-16 shows high accuracy in the field of face detection with over 95%. It's consisted of convolution layers, ReLU layers, max polling layers, and two fully connected layers followed by SoftMax for the output [33].

Resnet50 is convolutional neural network consist of 50 layers deep trained on million images from ImageNet can classify image into 1000 object, it has an input image size 224 by 224 this model score a high accuracy in the field of image classification 94% to classify 42different type of navigation marks in the sea and trained on 10260 navigation mark image [34]. Figure 8 shows the architectures of Resnet 50 convolutional neural network model.

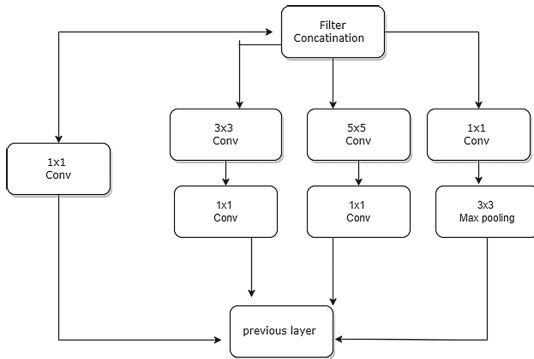


Fig. 6. Inception convolution neural network.

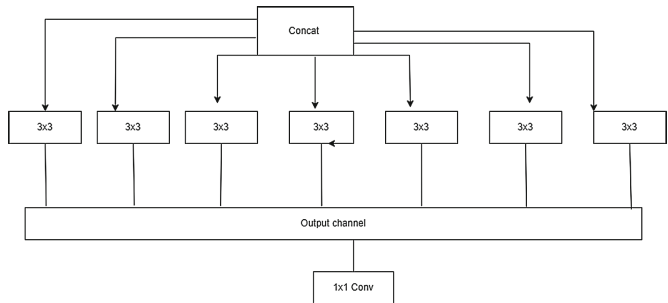


Fig. 7. Xception convolution neural network.

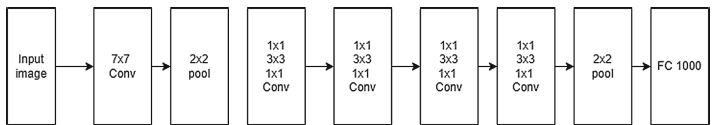


Fig. 8. Resnet 50 convolutional neural network.

NASNet-Large: The convolutional neural network NASNet-Large is trained on over a million images from the ImageNet database. The network can classify images into over 1000 different object categories, including keyboards, mice, pencils, and a variety of animals. As a result, the network has picked up rich feature representations for a variety of images. The image input size for the network is 331 by 331 pixels. To classify new images, NASNet-large is used instead of Google Net.

Resnet18 is a highly successful model architecture that earned first place in the ILSVRC 2015 classification competition. ResNet-18 is a convolutional neural network that is 18 layers deep. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. Resnet18 research score 93.8% accuracy in classifying it to cancer and no cancer.

4 Experimental Results

The pre-trained models above are trained and tested using the HAM10000 dataset. Data has been segmented into cancer and no cancer, which is a benchmark public database for machine learning. The data was divided into 80:20 Training: Testing ratios. And parameters have been changed to conclude the highest Accuracy. Parameters that have been changed are first of all Layers have been tried on 5 layers, pixelRange has been tried on $[-21\ 21]$, scaleRange has been tried on $[0.8\ 1.08]$, MiniBatchSize has been tried on 32, MaxEpochs have been tried on 5 and InitialLearnRate has been tried $1e-2$ and The models have been tried are Vgg16 and the accuracy is 90.6%, GoogleNet and the accuracy is 94.96%, Mobilenet v2 and the accuracy is 95.5%, Nasnet mobile and the accuracy is 93.11%, Squeezenet and the accuracy is 92.9%, Darknet19 and the accuracy is 90.16%, Xception and the accuracy is 96.66%, Shufflenet and the accuracy is 91.61%, Inception v3 and the accuracy is 95.76% and resnet50 and the accuracy is 95.66%, resnet18 is 93.8%,and NaseNet-large and the accuracy is 92.9% (Table 1).

Table 1. Metric values of pre-trained deep learning classifiers.

Model	Accuracy	Model	Accuracy	Model	Accuracy
Vgg16	90.6%	Squeezenet	92.9%	Inception V3	95.76%
GoogleNet	94.96%	Darknet19	90.16%	RESENET50	95.66%
Mobilenet v2	95.5%	Xception	96.66%	NASNETLARGE	92.9%
Nasnet mobile	93.11%	Shufflenet	91.61%	RESNET18	93.8%

5 Conclusion

Skin cancer drew the attention of the world in terms of its seriousness. Which led the recent researchers to work automated classifier system. This paper demonstrated the effectiveness of deep learning in automatic thermoscopic skin cancer classification with the Xception model trained on 10015 dermoscopy images from the HAM10000 dataset, which its Accuracy is reaching 96.66% during classification between cancer and no cancer. It can be concluded from the results that patients and physicians can effectively use the proposed system to diagnose skin cancer more accurately. This tool is more useful for rural areas where experts in the medical field may not be available. Since the tool is made more user-friendly and robust for images acquired in any conditions, it can serve the purpose of automatic diagnostics of Skin Cancer.

References

1. King, G., Zeng, L.: Replication data for: when can history be our guide? The pitfalls of counterfactual inference. Harv. Dataverse (2006)
2. Hanmer, M.J., Banks, A.J., White, I.K.: Replication data for: experiments to reduce the over-reporting of voting: a pipeline to the truth. Harv. Dataverse (2013)
3. Young, G.O.: Synthetic structure of industrial plastics. In: Peters, J. (ed.) *Plastics*, vol. 3, 2nd edn., pp. 15–64. McGraw-Hill, New York, NY, USA (1964)
4. Chen, W.-K.: *Linear Networks and Systems*, pp. 123–135. Wadsworth, Belmont, CA, USA (1993)
5. Ferlay, J., et al.: Cancer statistics for the year 2020: an overview. *Int. J. Cancer* (2021)
6. Apalla, Z., Nashan, D., Weller, R.B., Castellsaque, X.: Skin cancer: epidemiology, disease burden, pathophysiology, diagnosis and therapeutic approaches. *Dermatol. Ther.* **7**(Suppl. 1), 5–19 (2017)
7. Mader, K.S.: Skin cancer MNIST: HAM10000. Kaggle. <https://www.kaggle.com/kmader/skin-cancer-mnist-ham10000>. Accessed 2020
8. Skin Cancer Facts & Statistics [Internet]. The Skin Cancer Foundation. <https://www.skincancer.org/skincancer-information/skin-cancer-facts/>. Accessed 22 June 2021
9. Gavrilov, D., Lazarenko, L., Zakirov, E.: AI recognition in skin pathologies detection. In: 2019 International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI) (2019)
10. Sabri, M.A., Filali, Y., El Khoukhi, H., Aarab, A.: Skin cancer diagnosis using an improved ensemble machine learning model. In: 2020 International Conference on Intelligent Systems and Computer Vision (ISCV) (2020)
11. Maglogiannis, I., Doukas, C.N.: Overview of advanced computer vision systems for skin lesions characterization. *IEEE Trans. Inf. Technol. Biomed.* **13**(5), 721–733 (2009)
12. Hagerty, J., et al.: Deep learning and handcrafted method fusion: higher diagnostic accuracy for melanoma dermoscopy images. *IEEE J. Biomed. Health Inform.* **23**(4), 1385–1391 (2019)
13. Hekler, A., et al.: Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *Eur. J. Cancer* **118**, 91–96 (2019). <https://doi.org/10.1016/j.ejca.2019.06.012>
14. Davis, L.E., Shalin, S.C., Tackett, A.J.: Current state of melanoma diagnosis and treatment. *Cancer Biol. Ther.* **20**, 1366–1379 (2019)
15. Bohr, A., Memarzadeh, K.: The rise of artificial intelligence in healthcare applications. *Artif. Intell. Healthc.*, 25–60 (2020)
16. Pham, T.C., Luong, C.M., Visani, M., Hoang, V.D.: Deep CNN and data augmentation for skin lesion classification. In: Nguyen, N., Hoang, D., Hong, T.P., Pham, H., Trawiński, B. (eds.) *ACIIDS 2018. LNCS*, vol. 10752, pp. 573–582. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75420-8_54
17. Codella, N.C.F., et al.: Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM J. Res. Dev.* (2017)
18. Aruhan: A medical support application for public based on convolutional neural network to detect skin cancer. In: 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), pp. 253–257 (2021). <https://doi.org/10.1109/CEI52496.2021.9574496>
19. Zhou, D.-X.: Universality of deep convolutional neural networks. *Appl. Comput. Harmon. Anal.* **48**(2), 787–794 (2020)
20. Li, K.M., Li, E.C.: Skin lesion analysis towards melanoma detection via end-to-end deep learning of convolutional neural networks. arXiv preprint [arXiv:1807.08332](https://arxiv.org/abs/1807.08332) (2018)

21. Jaikishore, C.N., Udutalapally, V., Das, D.: AI driven edge device for screening skin lesion and its severity in peripheral communities. In: 2021 IEEE 18th India Council International Conference (INDICON) (2021)
22. Das, K., et al.: machine learning and its application in skin cancer. *Int. J. Environ. Res. Public Health* **18**, 13409 (2021)
23. LeCun, Y., et al.: Deep learning. *Nature* **521**(7553), 436–44 (2015). <https://doi.org/10.1038/nature14539>
24. Russakovsky, O. et al.: Imagenet large scale visual recognition challenge. *IJCV* (2015)
25. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258 (2017)
26. Dan, B., Sun, X., Liu, L.: Diseases and pests identification of *Lycium barbarum* using SE-MobileNet V2 algorithm. In: 2019 12th International Symposium on Computational Intelligence and Design (ISCID), , pp. 121–125 (2019)
27. Çakmak, M., Tenekeci, M.E.: Melanoma detection from dermoscopy images using Nas-net mobile with transfer learning. In: 2021 29th Signal Processing and Communications Applications Conference (SIU), pp. 1–4 (2021)
28. Park, E., Cui, X., Nguyen, T.H.B., Kim, H.: Presentation attack detection using a tiny fully convolutional network. *IEEE Trans. Inf. Forensics Secur.* **14**(11), 3016–3025 (2019). <https://doi.org/10.1109/TIFS.2019.2907184>
29. Setiawan, W., Purnama, A.: Tobacco leaf images clustering using DarkNet19 and K-means. In: 2020 6th Information Technology International Seminar (ITIS), pp. 269–273 (2020)
30. Li, Y., Lv, C.: SS-YOLO: an object detection algorithm based on YOLOv3 and ShuffleNet. In: 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 769–772 (2020)
31. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826 (2016)
32. Al Husaini, M.A.S., Habaei, M.H., Gunawan, T.S., Islam, M.R., Hameed, S.A.: Automatic breast cancer detection using inception V3 in thermography. In: 2021 8th International Conference on Computer and Communication Engineering (ICCCE), pp. 255–258 (2021)
33. Moid, M.A., Ajay Chaurasia, M.: Transfer learning-based plant disease detection and diagnosis system using Xception. In: 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 1–5 (2021)
34. Aung, H., Bobkov, A.V., Tun, N.L.: Face detection in real time live video using yolo algorithm based on VGG16 convolutional neural network. In: 2021 International Conference on Industrial Engineering, Applications (2021)



Predicted Phase Using Deep Neural Networks to Enhance Esophageal Speech

Madiha Amarjouf¹✉, Fadoua Bahja², Joseph Di-Martino³,
Mouhcine Chami¹, and El Hassan Ibn-Elhaj¹

¹ Research laboratory in Telecommunications Systems: Networks and Services (STRS), Research Team: Multimedia, Signal and Communications Systems (MUSICS), National Institute of Posts and Telecommunications (INPT),
Av. Allal Al Fassi, Rabat, Morocco

{amarjouf.madiha,chami,ibnelhaj}@inpt.ac.ma

² Laboratory of Innovation in Management and Engineering for Enterprise (LIMIE),
Institut Supérieur d'Ingénierie et des Affaires (ISGA Rabat), 27 Avenue Oqba, Agdal,
Rabat, Morocco

³ Loria - Laboratoire Lorrain de Recherche en Informatique et ses Applications, B.P.
239 54506 Vandœuvre-lès-Nancy, France
joseph.di-martino@loria.fr

Abstract. It is well known that the intelligibility and comprehension of pathological voices are notoriously poor, and since esophageal speech is one of those voices, the main idea behind our approach is to use Deep Neural Networks to enhance this speech. For that reason, the source phase and the target phase was aligned using the Fast Dynamic Time Warping method (FastDTW), in order to train the neural network. To retain the source speaker identity, the predicted phase along with source cepstral coefficients were used to reconstruct the output signal. The results obtained by this approach gave good performances.

Keywords: Esophageal speech · Phase · Speech enhancement · Spectrum · Cepstrum · Fast Fourier Transform · Deep neural networks · Fast dynamic time warping

1 Introduction

Esophageal speech is produced by people who underwent a total laryngectomy, which is a surgical operation that requires removing the larynx, because of an incident or advanced phases of laryngeal or hypopharyngeal cancer [1,2], the larynx, and more specifically the vocal folds are essential for producing speech [1–3]. This procedure impacts harshly respiration and phonation [1,2] due to the separation of the nasal cavity and the vocal tract [3]. For that reason, speech sounds cannot be produced normally [1]. Hence, laryngectomees could speak again without vocal folds [3,4], after voice rehabilitation [2], where they learn how to produce alaryngeal speech, using the vibrations of the pharyngoesophageal

part [5] with the assistance of a speech therapist [3]. This rehabilitation provides alternative ways for laryngectomees to speak, such as Esophageal Speech (ES), Tracheoesophageal Speech (TES) and Electrolarynx (EL Speech) [2,3]. To produce esophageal speech, laryngectomees need to aspire air through the stoma, which is a hole, then inject air into the esophagus [3–6]. Figure 1 shows this process. Esophageal speech is the main substitute way used by laryngectomees; it allows them to speak without the need of using any kind of device or equipment, not even using their hands [4]. It is characterized by restricted pitch range and intensity [6]. Nevertheless, its weak intelligibility makes it difficult to understand, because of the presence of unwanted noises related to its production process, the thing that impacts the social life of the speakers [1–5]. Thus, enhancing this speech is the aim of this study through predicted phase using Deep Neural Networks.

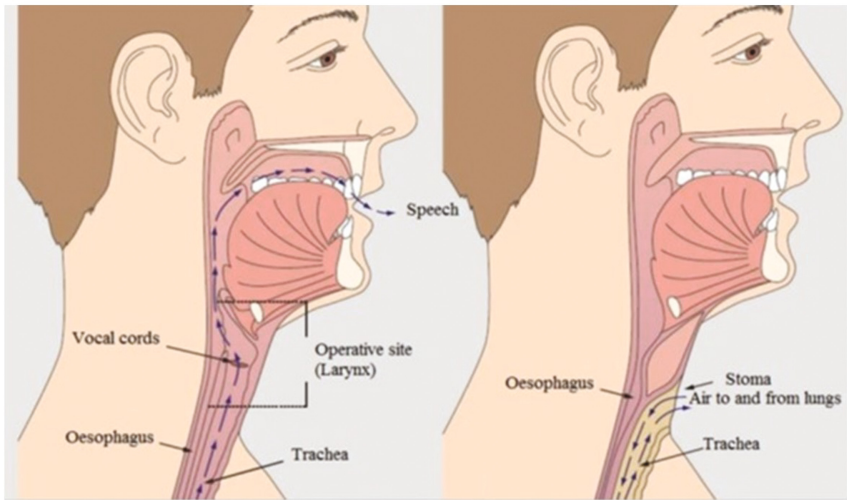


Fig. 1. The process of aspiring air and injecting it through stoma [8].

To improve esophageal speech, an enhancement aid was implemented to run in real-time [6]. The device was assessed using LPC, based on a formant analysis synthesis approach in which they replaced the voicing sources of esophageal speech with other voicing sources, generated using extracted inverse-filtered signals of ordinary utterers. As a result, the quality of esophageal speech was notably improved. Then, intelligibility and naturalness were remarkably improved by a new approach to enhancing esophageal speech, based on statistical voice conversion techniques [4]. The Esophageal-Speech-to-Speech conversion method preserved the linguistic information while transforming the esophageal speech into normal speech in a probabilistic way. Also, an effective voice quality control was proposed [1] to automatically manage voice quality in esophageal speech enhancement, based on the statistical esophageal Speech-to-Speech conversion method,

using eigenvoices and regression techniques. The multiple regression Gaussian mixture model and the kernel regression GMM were used to reach the manual manageability of the voice quality, of the transformed output. Even with few or no existing target samples, this approach was able to define converted voice quality. Another neuromimetic statistical conversion approach was applied to ES to enhance its naturalness and intelligibility by separately estimating the vocal tract and excitation cepstral coefficients [7]. Over and beyond the other conversion techniques, the source vocal tract was used to train the DNN model and the GMMs, while the cepstral excitation and phase was predicted by searching the target training space using a KD-tree. A DNN and GMM-based voice conversion system was proposed in order to improve the quality of esophageal speech [8]. For that, they tried to minimize the speech noises and retain the esophageal speaker identity, by applying the time-dilated Fourier cepstra to the source vocal tract. After that, a DNN was used as a nonlinear mapping function to train the dilated vocal tract. GMMs were the main voice conversion method. The excitation and phase were separately predicted as well, using a frame selection algorithm. For the purpose of making ES more intelligible, a novel DTW-free and parallel voice conversion approach were presented in [5]. Instead of using corresponding aligned normal speech as a target, this system used artificial speech, already matched in duration with the source ES. The voice conversion was implemented using the BLSTM network, and a set of Automatic Speech Recognition Systems were used as evaluation metrics. Enhancing esophageal speech through two hybrid denoising methods consisting of the wavelet-based methods was done in [9]. The wavelet-based methods were combined with the Wiener filter and the Time Dilated Fourier Cepstra, to denoise the vocal tract cepstrum in the first experiment, and to denoise the signal at the synthesis stage in the second experiment. Since the phase of a signal carries much relevant information, such as intelligibility, it is logical to investigate recovering some, if not all, of the magnitude information from the phase [10]. That is where came the aim of the current work, which is enhancing esophageal speech, by predicting the phase, after training the deep neural network (DNN) by the aligned source and target phases using FastDTW. The rest of this paper is structured as follows: The second part introduces the material and methods used in this work. Section three describes the results and discussion. Finally, the conclusion is given in the fourth section.

2 Material and Methods

2.1 Calculating the Phase

Initially, the spectrum of the signal is generated by applying the Fast Fourier Transform to the windowed signal, which is the product of the short-time audio signals multiplied by the normalized Hann window [11]. Thus, to compute the phase of the signal, an arc tangent moderator was applied to the real and imaginary parts of the spectrum. Figure 2 resume this process. The equation for calculating the phase is given by formula 1.



Fig. 2. The process of calculating the phase.

$$\phi = \arctan \frac{\Im(ft)}{\Re(ft)} \quad (1)$$

where $\Im(ft)$ designs the imaginary part of “the spectrum of the signal”, and $\Re(ft)$ represents its real part.

2.2 Alignment

To determine the best alignment between two time series, Dynamic Time Warping (DTW) is usually used. This method is frequently used in speech processing to assess if the two speech signals reflect the same uttered word. Although the time between sounds and the length of each uttered sound may differ, the overall shape of the speech wave-forms must be consistent [12]. Unlike the standard dynamic time warping algorithms, FastDTW [12] which is capable of determining an exact estimate of the best warp path between two time series, employs a multilevel approach that iteratively projects and refines a solution from a coarse resolution. By employing this approach, the FastDTW algorithm bypasses the typical DTW algorithm’s brute-force dynamic programming technique. Fast-DTW was adopted to align the two extracted phase sequences of both the source and the target.

2.3 The Deep Neural Network Model

In the enhancement model, a four architecture layer DNN was used to train the phase source and target inputs. Figure 3 shows the flow chart of this architecture, where X_p denotes the input phase, p is the length of the phase sequence and N is the number of neurons used in the three deep layers.

2.4 The Proposed Enhancement Model

The proposed model aligned the source phase sequence $[S_0 i...S_{N-1} i]$ and the target phase sequence $[T_0 i...T_{N-1} i]$ using the FastDTW, yielding the best calculated path of the source and target phases. The component associated with the path’s indices was then used as the X_{train} and Y_{train} training inputs. The training inputs were fed into a sequential, four-layer DNN; the three deep layers have 512 neurons, and the input layer and output layer have 256 neurons. All four layers were of type dense, which is the standard deeply connected neural network layer in which each neuron receives input from all neurons of the previous layer. The activation function used for each layer was the ReLu function, which stands for

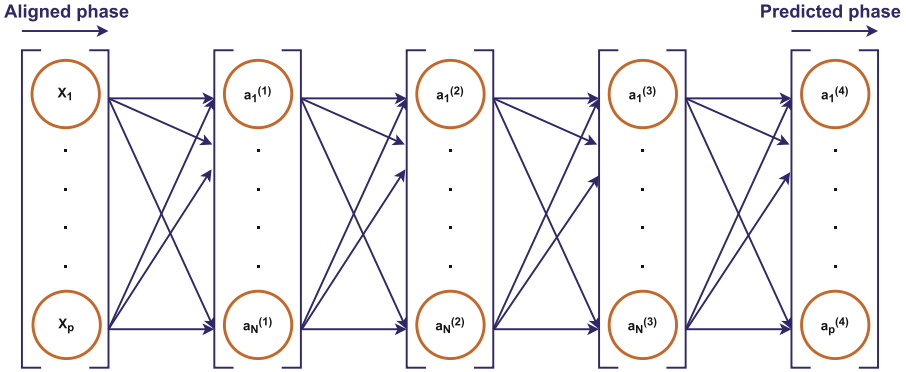


Fig. 3. The architecture of the neural network used.

the Rectified Linear Unit. The predicted phase along with the extracted cepstral features of the esophageal speaker were used to reconstruct the speech signal. The proposed model of phase prediction approach is shown in Fig. 4.

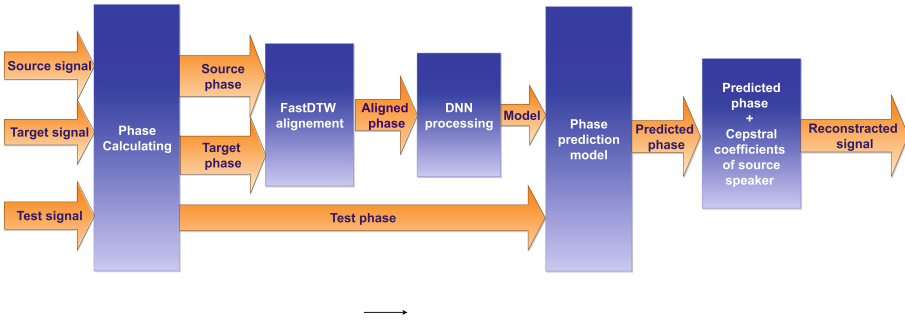


Fig. 4. The proposed model of phase prediction approach.

2.5 Dataset

A French database spoken by French male speakers and containing both esophageal and laryngeal corpora was used in this work. “PC” is the esophageal corpus, while “AL” is the parallel normal corpus. Each wave file of the two corpora contains phonetically balanced phrases. 150 esophageal wave files (PC speaker) were used as source files, and 150 normal wave files (AL speaker) were used as target files. 10% of the training set was used as a validation test, while 10 files (not from the training set) were used for the test. Table 1 gathers the experimental parameters used.

Table 1. Experimental parameters.

Sampling rate	16 kHz
Step size	4 ms
Analysis window size	512
Number of training wave files	150
Number of validation wave files	15
Number of test wave files	10
Phase length for inputs and outputs	256
DNN structure	512 * 3
Epochs	30

3 Results and Discussion

In order to evaluate the results, the Signal-to-Error Ratio (SER) and the Cepstral Distance (CD) were calculated. The formulas for SER and CD were given by [7] as follows:

$$SER = -10 * \log_{10} \frac{\sum_k \|y_k - \hat{y}_k\|^2}{\sum_k \|y_k\|^2} \quad (2)$$

$$CD = \frac{10}{\log_{10} \sqrt{\sum_k (\hat{y}_k - y_k)^2}} \quad (3)$$

where y_k and $\hat{y}_k$ are the target and converted cepstral vectors, respectively.

Table 2 and Table 3 summarize the results of the average of the SER and the CD of the 10 test files.

Table 2. SER and CD of PC test wave files.

	File 1	File 2	File 3	File 4	File 5	File 6	File 7	File 8	File 9	File 10
SER [db]	4.1237	3.7008	3.6869	3.8393	3.7183	3.5802	4.2799	2.9279	2.6764	3.0079
CD [db]	6.9654	7.0694	6.7382	7.0353	7.0064	7.0155	7.0638	6.8432	6.7358	6.8113

Table 3. The mean of the above SER and CD.

Average SER	Average CD
3.5541	6.9284

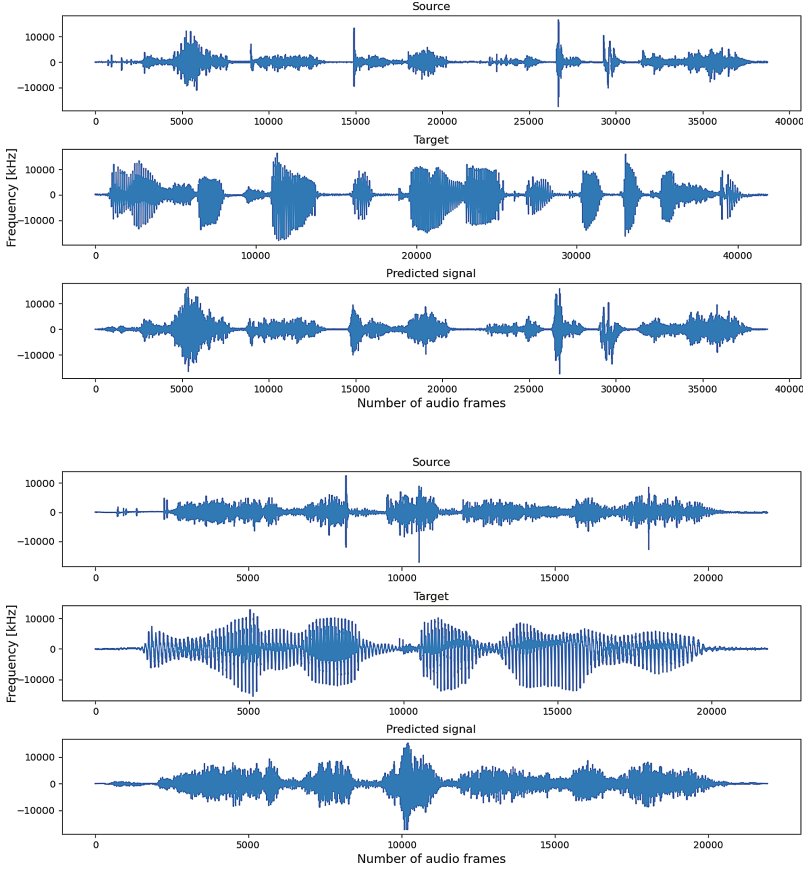


Fig. 5. The signals of two different sentences, each sentence has three wave files of the same spoken speech corresponding to utterance of the source, the target and the reconstructed signal from the predicted phase.

As demonstrated by Table 2, the Cepstral Distance values of the test files are almost identical, which proves that the predicted phase model has the same effect on the test files and demonstrates that the identity of the speaker was preserved. The Signal-to-Error-Ratio rates show that the signals still have some noise, which explains the low SER rates. Also, Fig. 5 presents the signals of two different sentences. Each sentence is represented by three speech signal corresponding respectively to the source, the target, and the reconstructed signal from the predicted phase. As shown in this figure, we can notice that the shape of the predicted signals for both sentences are different than the shape of the signals of the source speech files and slightly similar to the target wave files.

4 Conclusion

A predicted phase approach to improving esophageal speech was presented in this paper. From the proposed enhancement model, we can conclude that using FastDTW for aligning the source and target phases has shown an improvement in the predicted results. Furthermore, the reconstructed signal with the predicted phase and the same cepstral features as the source speaker improved the esophageal speech while retaining the esophageal speaker identity. In this paper, we demonstrate that using only phase to enhance esophageal speech is possible and gives important results. However, the outcomes of this approach need to be improved in future work using additional methods to reduce the noise of the speech signals.

References

1. Yamamoto, K., Toda, T., Doi, H., Saruwatari, H., Shikano K.: Statistical approach to voice quality control in esophageal speech enhancement. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4497-4500. IEEE, Kyoto (2012). <https://doi.org/10.1109/ICASSP.2012.6287949>
2. Ouattassi, N., et al.: Acoustic assessment of eryngomphonic speech of Moroccan laryngectomized patients. *Pan Afr. Med. J.* **21**, 270 (2015). <https://doi.org/10.11604/pamj.2015.21.270.4301>
3. García, S.L., Raman, S., Hernández, R.I., Navas, C.E., Sanchez, J., Saratxaga, I.: A Spanish multispeaker database of esophageal speech. *Comput. Speech Lang.* **66** (2021). <https://doi.org/10.1016/j.csl.2020.101168>
4. Doi, H., Nakamura, K., Toda, T., Saruwatari, H., Shikano, K.: Esophageal speech enhancement based on statistical voice conversion with Gaussian mixture models. *IEICE Trans. Inf. Syst.* **E93-D**(9), 2472–2482. (2010). <https://doi.org/10.1007/1234567890>
5. Raman, S., Sarasola, X., Navas, E., Hernaez, I.: Enrichment of oesophageal speech: voice conversion with duration-matched synthetic speech as target. *Appl. Sci.* **11**, 5940 (2021). <https://doi.org/10.3390/app11135940>
6. Matsui, K., Hara, N.: Enhancement of esophageal speech using formant synthesis. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258), vol. 1, pp. 81–84 (1999). <https://doi.org/10.1109/ICASSP.1999.758067>
7. Ben Othmane, I., Di Martino, J., Ouni, K.: Enhancement of esophageal speech using statistical and neuromimetic voice conversion techniques. *J. Int. Sci. Gen. Appl.* **1**(1), 10. {hal-01724375} (2018)
8. Ben Othmane, I., Di Martino, J., Ouni, K.: Enhancement of esophageal speech obtained by a voice conversion technique using time dilated Fourier cepstra. *Int. J. Speech Technol.* **22**(1), 99–110 (2018). <https://doi.org/10.1007/s10772-018-09579-1>
9. Amarjouf, M., Bahja, F., Di Martino, J., Chami, M., Ibn Elhaj El, H.: Denoising esophageal speech using combination of complex and discrete wavelet transform with wiener filter and time dilated Fourier Cepstra. In: ITM Web Conference The 4th International Conference on Computing and Wireless Communication Systems (ICWCWS 2022), vol. 48 (2022). <https://doi.org/10.1051/itmconf/20224803004>

10. Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. *Proc. IEEE* **69**(5), 529–541 (1981). <https://doi.org/10.1109/PROC.1981.12022>
11. Griffin, D., Lim, J.: Signal estimation from modified short-time Fourier transform. *IEEE Trans. Acoust. Speech Sig. Process.* **32**(2), 236–243 (1984). <https://doi.org/10.1109/TASSP.1984.1164317>
12. Salvador, S., Chan, P.: FastDTW: toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* **11**(5), 561–580 (2007). <https://doi.org/10.3233/IDA-2007-11508>



State of the Art Literature on Anti-money Laundering Using Machine Learning and Deep Learning Techniques

Bekach Youssef¹(✉), Frikh Bouchra¹, and Ouhbi Brahim²

¹ LIASSE Lab ENSA, University Sidi Mohamed Ben Abdellah, Fes, Morocco
{youssef.bekach,bouchra.frikh}@usmba.ac.ma

² Mathematical and Informatics Modeling Laboratory, High School of Arts and Crafts-Meknès,
Meknes, Morocco
b.ouhbi@umi.ac.ma

Abstract. Money laundering (ML) is the process by which criminals convert illicit funds gained from drug sales, human trafficking, and other illegal operations into legal funds that can later be spent freely and without restrictions. The annual amounts of money involved in ML are in the billions of dollars, and it is believed that it accounts for 2% to 5% of worldwide GDP [10]. Governments around the world are trying to combat this type of criminal activity by requiring banks and other financial institutions to report any such activity. In this paper, we will conduct an in-depth examination of the current literature on anti-money laundering, focusing on the areas of machine learning, deep learning, data mining, and big data. After reviewing the papers, we noticed that while various graph algorithms are being used, such as graph embedding and centrality algorithms, there is a shortage of community detection algorithms that could help identifying organizations and groups involved in money laundering. This gap in the literature suggests a potential area for future research to investigate the use of community detection algorithms in anti-money laundering systems to enhance their ability to detect illegal activity.

Keywords: Anti-money laundering · Machine learning · Deep learning · Data mining

1 Introduction

Criminals use money laundering to disguise and divert the illegally obtained proceeds so that they appear to come from legitimate sources. The money laundering process is divided into three stages. The first stage is placement, followed by layering, and finally integration. During the placement stage, the launderer deposits illegal funds in local and international financial institutions, then makes as many money transfers as possible through a complicated network of accounts to make tracing the source account difficult in the layering stage. Finally, during the integration stage, the funds are re-used in the legitimate economy.

Most banks and financial institutions rely on traditional rule-based systems developed by domain experts to combat this type of financial crime. These systems send alerts about suspicious transactions based on the amount of the transaction, the countries involved, and a variety of other factors. However, it has several disadvantages in that it necessitates a significant amount of human effort as the investigators must review each of these transactions to determine whether they should be blocked or released. Also, the false positive rate can exceed 90% of total cases [4], and these rules should be updated on a regular basis to keep up with new fraud schemes used by criminals. Therefore, finding more intelligent solutions to this problem has become necessary.

Artificial intelligence appears to be a promising solution to this problem, as it has already demonstrated its capabilities in a variety of fields, including natural language processing, computer vision, healthcare, finance, etc. Several artificial intelligence anti-money laundering systems have already been developed, and we will attempt to review the most recent works in this paper.

The rest of the paper is structured as follows. The research methodology is described in Sect. 2. Section 3 summarizes precedent anti money laundering review papers. Section 4 includes studies that have been reviewed in this paper. Finally, Sect. 5 concludes this work.

2 Research Methodology

The purpose of this study was to conduct a systematic search of the latest anti-money laundering techniques and their applications using the keywords “Anti-money laundering,” “money laundering,” “machine learning,” “deep learning,” “data mining,” and “big data” through databases like Science Direct, Ieeexplore, and ResearchGate and the website “scholar.google.com” from 2017 to 2023.

Based on the titles and abstracts and the year of publication, only papers that were relevant to anti-money laundering were chosen, which are 4 review papers (see Sect. 3) and 19 journal and conference papers (see Sect. 4).

3 Literature Survey

A critical review of anti-money laundering systems was conducted by [1], focusing on deep learning methods. They found that models such as the convolutional neural network (CNN), variant graph based CNN, scalable CNN, and multi-channel CNN based on natural language processing (NLP), the AutoEncoder, and the multi-layer perceptron (MLP) were most commonly used. They also investigated the interpretability of machine learning systems in this field, discovering that 51% of these systems were uninterpretable with no explainable artificial intelligence (XAI) methods. Based on this, they suggested that future research should concentrate on using explainable AI approaches, link analysis methods such as graph mining and social network analysis, ensemble learning, and unsupervised learning methods. They also discovered that the majority of works in this field were evaluated using either a small data set or a synthetic dataset, so they advised concentrating on larger real-time transaction data and data preprocessing step.

A survey of data mining techniques for anti-money laundering has been presented in [2], with the majority of papers using clustering techniques such as k means, CLOPE algorithm, and minimum spanning tree, as well as rule-based methods such as Bayesian networks, Ontology Semantic Web, Radial basis function network (RBF), support vector machine (SVM), and decision tree (DT). They also presented methods based on social network analysis.

A systematic review of detecting illicit transactions systems, including anti-money laundering systems in the field of Bitcoin, was conducted by [18]. 25 papers were presented and categorized into three groups: topological analysis studies based primarily on graph algorithms; unsupervised learning methods like clustering, principal component analysis (PCA), and isolated forests; and supervised learning methods like random forests, boosting, and neural networks.

[20] presents a review of artificial intelligence systems for anti-money laundering. Numerous studies were examined, including link analysis and network analysis systems that rely on centrality algorithms to prevent money laundering and discover connections inside money laundering gangs. They also presented risk classification studies that used machine learning to assign a risk score to the transactions. In addition, they present their own system for detecting money laundering using NLP and deep learning methods, including sentiment analysis, named entity recognition, relation extraction, and other methods based on models such as CNN and long short term memory (LSTM). They claimed their system could reduce the time and cost of investigations by approximately 30%.

4 State of the Art AML Using Machine Learning and Deep Learning

4.1 Supervised Methods

Supervised learning algorithms can be categorized into two categories: classification and regression. On the one hand, the classification algorithms are used to predict discrete values, and on the other hand, the regression algorithms are used to predict continuous values. Both of these algorithm types, however, require a labeled training set.

Many supervised methods have been created to detect money laundering, mostly as a binary classification problem between two classes [3, 4], but also as a multiclass problem [6], and sometimes as a regression problem [5].

As previously stated, most banks and financial institutions use rule-based anti-money laundering (AML) systems, which trigger alerts for suspicious transactions, which are then investigated by human experts. Most supervised methods in this field rely on the results of these investigations to label the training dataset. In [3], they suggest a system for future work where suspicious transactions are detected by a watch list filter system that employs a string matching algorithm to verify transaction information such as names, aka, addresses, and countries against preloaded blacklists. The red-flagged transactions are then inspected by investigators and labeled as either a money laundering case or a regular case. A machine learning component is then trained based on these labels in the first stage. The second stage is an advising phase, in which the machine learning

component will just act as an advisor and the investigators will make the final decision; the third stage is the take action phase, in which the model will make the decision without the need for investigators. In their experiment, they used decision tree (DT), support vector machine (SVM), and Naive Bayes (NB) algorithms on a dataset of 1500 blocked transactions, using attributes from the watch-list filtering, the Swift message information, and know your customer (KYC). The results showed that SVM outperformed the other algorithms, which got 85% in accuracy, precision, and recall.

In [19], they proposed a machine-learning-based anti-money laundering monitoring system by using two models chained one after the other. The first model is a logistic regression model that filters out genuine customers, and the second model is an extreme gradient boosting model that is used for the final classification of the remaining customers. Once per week, the model classifies the customers as licit or illicit based on two thresholds. The first is a base threshold used to classify the risk of a customer, and the second is a jump threshold used to compare the risk score of a customer in two consecutive weeks, where the idea is to avoid redundant alerts and ensure that alerts are only raised if there is both a risk of illegality and a major change in a customer's behavior. In order to accurately measure the performance of this periodic AML monitoring system, they proposed new custom metrics that take into account the temporal dimension, mainly a custom recall (43%), and a custom precision (40%). They also addressed the interpretability of the model by using Shapley values of the features, which allows them to extract features that provided the highest contribution in the classification task. Using a real-life bank transaction dataset, they claimed that their custom monitoring system was able to reduce the number of redundant alerts per customer in comparison with a default monitoring system, while the alerts generated besides the explained features generated by the Shapley values were relevant according to the investigators.

Launderers and financial criminals always try to innovate in their crime methods, so feature engineering can play a big role in generating new features that can detect these crime patterns. In [4], for example, they present a system to detect money laundering transactions based on three types of data: bank transaction data, customer properties-related CRM data, and a novel set of features that they create using time frequency analysis, such as mean, variance, Kurtosis, and skewness. The data labels were generated from previously reported suspicious activities. A random forest algorithm was used to evaluate the transactions. The results prove that by using the new set of features besides the traditional attributes, the performance of the model was increased, reaching 92% in ROC curve AUC.

In [15], a new set of features were used in addition to the original Scotiabank customer dataset features, such as monthly averages and yearly sums of incoming and outgoing transaction amounts for each account, as well as proportions of total yearly transactions (i.e., the year's sum of incoming debit divided by the year's total amount of incoming funds). Customers were divided into three groups by experts: low-risk, medium-risk, and high-risk customers. Several supervised techniques, including logistic regression with LASSO regularization, traditional logistic regression, k-nearest neighbors, and extreme gradient boosting, were trained using 70% of the dataset. The results showed that the new set of features improved the performance of the models with a 0.79 F1 score, 0.72 accuracy, 0.71 sensitivity, and 0.74 specificity obtained by the logistic regression.

In [5], they used a regression tree model to quantify the money laundering inherent risk associated with a customer profile, using a dataset from a real estate financial institution that contains five independent variables (legal entity, the origin of the customer, the economic activity, the seniority, and the contracted product). The data was labeled by experts based on criteria given by Mexico's Ministry of Finance and Public Credit, and statistical tests revealed that the variables "contracted product" and "customer seniority" in the institution are the only ones that are statistically significant. The other independent variables were left out of the regression tree's construction. The regression tree classified the customers as high-risk or low-risk based on a probability threshold of 0.5. The results showed that 78.7% were correctly predicted as low-risk profiles, and 100% were correctly predicted as high-risk profiles.

In [6], they developed a supervised machine learning method based on the extreme gradient boosting algorithm to detect suspicious money laundering transactions using a dataset from Norway's largest bank, DNB, where they had four classes of labels. The first class label (A) is for no-alert transactions (normal transactions). The second-class label (B) is for alerted transactions that turn out to be non-money laundering cases after the first examination by the investigators. The third class label (C) is for the remaining alerted transactions that turn out to be non-money laundering cases after the second thorough investigation by experienced inspectors. The fourth-class label (D) is for the reported transactions after the examination. These labels were used to train three models. The first is a binary model with reported class transactions (D) and non-reported class transactions (A + B + C), the second a multiclass model with four classes (A), (B), (C), and (D), and the third a multiclass model with three classes (A), (D), and (B + C). And besides the AUC and the Brier score, they introduced a novel evaluation metric which is the proportion of positive predictions when adjusting the classification threshold such that the true positive rate (TPR) is at a certain level (g) referred to as PPP ($TPR = g$). The results showed that both multiclass models had slightly better AUC (91%).

Despite the fact that money laundering is estimated to be worth billions of dollars every year, the number of cases discovered is typically small in comparison to the millions of transactions made every day, especially with the rise of e-commerce. This results in highly imbalanced transaction data, which can be difficult for supervised models to deal with. A trained model on this type of data will be biased towards the majority class prediction. Several techniques can be used to address this problem. For example, in [7], using real data from a U.S. financial institution, a label of "1" was assigned to the suspicious transactions that were reported as suspicious activity reports (SAR) by the investigators and "0" otherwise. And as the dataset is highly imbalanced, with only 0.71% of class "1" (SAR) in the training data and 0.28% in the test data, the authors used over-sampling and under-sampling techniques to balance the two classes, and by using several machine learning algorithms. The best results were obtained with the ANN using AUC and ordinary least squares regression analysis as evaluation metrics.

In [17], they proposed a deep learning approach to detect money laundering using a dataset from Spar Nord, a Danish bank, with 5 transaction-related features such as transaction monetary value and 14 client-related features extracted from the KYC information. Using feature transformation techniques such as ordinal encoding of categorical features, log transformation, standardization, and embedding, 22 transaction-related features and 15 client-related features were generated. The transaction-related features were then passed through a series of sequence processing blocks, primarily built with lstm,

gru, and transformer blocks, to create a new representation of these features. Finally, the new representation of the transaction features and the client-related features was used to train a two-layer feedforward network with an output neuron employing the sigmoid activation function. The obtained results were compared with several machine learning models, such as the gradient boosting model, the support vector machine model, and others. They found that their deep learning approach had better performance, with 91% of the area under a ROC curve (ROC AUC).

Aside from banks and financial institutions, the blockchain's emergence has opened a large door for criminals to exploit it in their illegal activities. Several methods were created to deal with ML in Cryptocurrency. In [8], for example, they used machine learning algorithms such as Logistic Regression (LR), Random Forest (RF), Multilayer Perceptron (MLP), and Graph Convolutional Networks (GCN) to detect money laundering on the Elliptic Data Set, which is a labelled time series graph of over 200K Bitcoin transactions (nodes), 234K directed payment flows (edges), and 166 node features. 70% was used for the training and the remaining 30% was used for the testing. The best results were obtained with Random Forest using elliptical dataset features along with node embedding features, with 0.971, 0.675, and 0.796 in accuracy, recall, and F1, respectively.

To improve on the results obtained by the GCN in [8], they proposed in [21] a system combining the GCN and a linear model with the goal of learning both graph and linear feature representation, and they managed to obtain better results. The same dataset was used in [14] as well, and four supervised learning algorithms, namely logistic regression, random forest, support vector machine, and decision tree, where random forest had the best results as well, using precision, recall, and F1-Score as evaluation metrics.

Several anomaly detection techniques, including isolation forest, local outlier factor, and others, as well as active learning techniques with four query strategies—uncertainty sampling, expected model change, the elliptic envelope, and the isolation forest—were studied in comparison to the supervised methods used in [8] in [23]. They discovered that anomaly detection methods perform noticeably worse than supervised methods, whereas they were able to achieve comparable or slightly better results than the supervised methods when using active learning techniques with just 5% of the total labeled samples.

In [22], they suggested a graph-learning-based money laundering detection system for bitcoin. They used a graph dataset they had constructed using the Bitcoin core client, and they extracted attributes including the timestamp, number, and value of each transaction's input and output. The dataset was categorized based on the transaction addresses: those arriving from money-laundering sites like AlphaBay and BTC-e were categorized as such, while those coming from ordinary bitcoin services were categorized as such. Four different types of features were used: node2vec embeddings, deep walk embeddings, immediate neighbors features, and local transaction features, to train an adaboost classifier with a decision tree-based estimator classifier. The results demonstrated that employing an ensemble of node2vec embedding and the local features as input generated the best results, where accuracy and the f1 score were used as evaluation metrics. (See more details in Table 1).

Table 1. Supervised methods for anti-money laundering

Article	Year	Dataset	Type of the data	Time window	Data partitioning	Data Balancing technique	Method	Evaluation metric
[3]	2021	Dataset contains 1500 transactions (746 blocked/754 released)	Real bank transactions data	–	Stratified kfold	–	Support vector machine-naïve Bays-decision tree	Confusion matrix, accuracy, precision, recall
[4]	2021	Akbank transaction and crm data For 6,680 customers	Real bank transactions data	6 months of data	Cross-validation + Test on 4,263 customers' s transactions in may, 2020)	–	Random forest	Confusion matrix, auc-fpr, Fmr, precision, Recall and f1-score
[5]	2020	Real estate financial institution dataset	Real data	–	–	–	Decision regression tree	Confusion matrix
[6]	2020	Dataset from the norway's largest bank, dnb	Real banking data	1 april 2014 to 31 december 2016	Training set transactions from 01/04/2014 to 30/06/2016, Test set:transactions from 1/07/2016 to 31/12/2016	–	Extreme gradient boosting	The brier score, the auc, proportion of positive predictions ppp (tpr = g)
[7]	2019	Data from a u.s. Financial institution	Real data	13 months	Training (8 months data)/testing (5 months data)	Under-sampling; over-sampling	Maximum likelihood logistic regression; Bayes logistic regression; Decision tree; Random forest; Support vector machine; Artificial neural network	Area under the curve (auc); Ordinary least squares (OLS)

(continued)

Table 1. (continued)

Article	Year	Dataset	Type of the data	Time window	Data partitioning	Data Balancing technique	Method	Evaluation metric
[8]	2019	The Elliptic Data Set	Real bitcoin transactions	–	70% training set; 30%test set	–	Multilayer Perceptron; Logistic Regression; Random Forest; Graph deep learning	Precision; Recall; F1 score
[14]	2021	The Elliptic Data Set	Real bitcoin transactions	–	90% training Set (Stratified 10-fold Cross Validation technique) 10% test set	–	Logistic regression, random forest, support vector machine, and decision tree	Precision; Recall; F1 score
[21]	2022	The Elliptic Data Set	Real bitcoin transactions	–	70% training set; 30%test set		Graph Convolutional Network Linear model	Precision; Recall; F1 score
[23]		The Elliptic Data Set	Real bitcoin transactions	–	70% training set; 30%test set		Logistic Regression; Random Forest; Xgboost	F1 score;
[22]	2019	Bitcoin dataset	Real bitcoin transactions	From 07/2014 to 05/2017	–	–	Adaboost	F1 score; Accuracy

(continued)

Table 1. (continued)

Article	Year	Dataset	Type of the data	Time window	Data partitioning	Data Balancing technique	Method	Evaluation metric
[15]	2021	Scotiabank, a large bank in Canada	Real transactions dataset	from April 15, 2019 to April 15, 2020	70% training set 30% training set	-	Logistic regression, logistic regression with LASSO regularization (LASSO), K-nearest neighbours (KNN), and extreme gradient boosted models (XGBoost)	Accuracy F1 score Recall Specificity
[17]	2023	Spar Nord, a Danish bank	Real transactions dataset	January 1, 2020, to January 31, 2022	46,750 transactions for training 7,259 transaction for validation	-	Gru, lstm, transformers, Feedforward layers	Area under a ROC curve
[19]	2022	Anonymous bank	Real transaction dataset	March 2018 to April 2020	Training set: transactions from 09/2018 to 10/2019 Testing set: transactions from 10/2019 to 04/2020	-	Logistic regression + extreme gradient boosting	Custom F1 score, custom precision, custom recall

Although these supervised methods appear to be effective at detecting money laundering, they have a number of flaws that should be addressed. One is the need for inspectors' intervention in order to create a labeled dataset, as not all banks and financial institutions have the ability or financial resources to hire them, and even if they do, the resulting dataset will always contain a very small number of money laundering cases, resulting in an imbalanced dataset. As a consequence, the majority of the studies we reviewed either used a small dataset [3], or a large unbalanced dataset that needed to be balanced [7]. To address this problem, numerous studies on unsupervised methods for detecting money laundering have been published. In the next section, we'll go over some of them.

4.2 Unsupervised Methods

Unsupervised learning, unlike supervised learning, does not require any labeled data, which makes it a perfect fit for the money laundering problem. Unsupervised learning algorithms include clustering, anomaly detection, network analysis, dimensionality reduction, and others. Some of these algorithms have been used in the papers that we will explore in this section.

In [9], they developed a network-based system to detect money laundering in a factoring company. They started by creating a transaction network between sellers and debtors that included 559 nodes (sellers/debtors and nodes with a double role) and 33670 links (representing money transfers in Italy and abroad), and based on risk factors reported by the Financial Intelligence Unit of the Bank of Italy, four new networks were built. The first is a transaction-based network (link-based network), where a range of weights are given to the links (1 to 3) based on the transaction amount. The higher the amount of the transaction, the higher the weight. The second and the third networks are node-based networks, depending on the geographical area and the economic sector of the nodes, where high and low-risk nodes were rated by two domain experts. The final network, called the Tacit Link Network, links a seller node with a debtor node when they have the same owner or representative. The purpose of this network is to detect money launderers who can create fake financial operations in their name. Risk profiles were labeled with "1" and normal profiles with "0" in the created networks, after examining the Italian court records to check whether they had been involved in anti-money laundering investigations before. Furthermore, by employing social network metrics such as "in-degree", "out-degree", "all-degree", "closeness", "betweenness centrality", and the "network constraint", they found that a larger "in-degree centrality" on the Geographical Area Network and a lower network "constraint" on the economic sector network increase the probability of dealing with a risk profile.

In [10], they present two unsupervised systems to detect anomalous money laundering transactions based on two anomaly detection algorithms: Isolation Forest and One Class support vector machine. The results showed that the Isolation Forest outperformed the One-Class SVM, detecting 11825 versus 11222 anomalies, respectively, where domain experts approved that all cases classified by the system are high-risk. Whereas 50 cases were clearly identified as suspects, with 414 suspicious transactions and 4 money-laundering groups.

In [11], they propose a method for detecting the full flow of money laundering from source to destination using a multipartite graph. Multipartite suspicious money laundering sub-graphs were extracted using a new anomalousness metric that took into consideration two main criteria: First, the density of the sub-graph as money launderers frequently conduct high volumes of transactions, resulting in a dense and high-volume subgraph. Second, the balance between the weighted indegree and outdegree of the middle accounts in the suspicious subgraph, where in the money laundering process, these accounts act only as a bridge between the source and destination accounts. The results showed that the effectiveness of their method for detecting money laundering outperforms the most recent methods using the F1 score and the areas under the F scale curve as evaluation metrics.

In [12], they presented dimensionality reduction methods such as AutoEncoder and Principal Component Analysis to detect anomalies in exporters of Brazilian exports, using a database of exports of goods and products that occurred in Brazil in 2014. Because it was 20 times faster than PCA, the AutoEncoder algorithm was chosen for testing. They used the mean squared error (MSE) to assess the model, which measures the distance between predictions and actual data, with the higher the MSE value, the more unusual the record. Experts confirmed some fraud cases they already knew about among the fifty companies with the highest MSE.

In [13], they introduce a three-phase system for detecting money laundering and financial terrorism based on dataset from a Mexican financial institution. In the first phase, experts used fuzzy logic to assign a risk metric (r) and a level of uncertainty (a) to each value of the variables based on FIU and FATF typologies for money laundering and terrorism financing crimes. Clusters were formed in the second phase using clustering methods like strict competitive learning, self-organizing maps, C-means, and neural gas. The best clustering technique was determined using the Calinski-Harabasz index (CH), and the c-means algorithm was found to be the most effective. In the final phase, an indicator of abnormality based on the variance of the variables is applied to each transaction in the riskiest clusters to detect any abnormal behavior that signals the presence of crime. The transaction is then classified as risky, and an alert is generated if the indicator passes a certain threshold. The results showed that their method was able to achieve good results using the confusion matrix, the balanced accuracy, and the balanced error rate as evaluation metrics.

In [16], they proposed a method for detecting money laundering using link analysis graphs. First, suspicious transactions were extracted from the original bank transaction data set using SQL queries, taking into account criteria developed by the FIU to identify potential money laundering patterns (e.g., u-turn cash flow, cash flows involving interesting countries, high volume of transactions within a short period, etc.). Then different graphs were generated based on these transactions using the Graphviz visualization engine and twopi graph algorithm, where the nodes represent the accounts and the edges represent the transactions. The goal of this method is to improve the investigator's ability to see and target suspicious patterns quickly (See more details in Table 2).

Money laundering is a serious problem that is often committed by organizations rather than just individuals. Some of the papers we reviewed attempted to tackle this problem and try to detect these groups by using graph algorithms to identify patterns and

Table 2. Unsupervised methods for anti-money laundering

Article	Year	Dataset	Type of the data	Time window	Method	Evaluation metric
[9]	2017	Central database of a medium-large factoring company in Italy	Real dataset	November 2013 to June 2015	Social network analysis	In-degree, out-degree, all-degree, closeness and betweenness centrality
[10]	2020	118250 transactions with three fraud patterns (Fan-in, Fan-out and Cycle)	Synthetic dataset from AMLSim	–	Isolation forest; one class SVM	–
[11]	2020	– CBank Dataset;	Real data	Aug/07/2017 – Aug/13/2017	Multipartite graph	Areas under curve of F-measure(FAUC); F1 score
		– Czech Financial Dataset (CFD)	Real anonymized Czech bank transactions	Jan/01/1993 – Dec/31/1998		
[12]	2017	Dataset of exports of goods and products that occurred in Brazil in 2014	Real data	Records from 2014	– Autoencoder – Principal component analysis	Mean square error
[13]	2021	Data warehouse of a Mexican institution	Real data	Training set from January 2020; Test set from February 2020	Clustering techniques: strict competitive learning, self-organizing-map, C-means, and neural gas	Confusion matrix; Balanced accuracy; Balanced error rate;
[16]	2019	Bank transactions	Real data	2014 to 2016	Data visualization	-

connections between transactions, such as centrality algorithms [9] or graph embedding algorithms [21]. However, there is an absence of community detection algorithms in these papers, which could also be used to detect money laundering gangs. Community detection algorithms are able to identify groups of individuals or entities that are closely connected and could potentially be used to uncover hidden networks involved in money laundering.

In the future, researchers can utilize these community detection algorithms to further enhance the detection of money laundering activities. By identifying these closely connected groups or communities, researchers can potentially uncover hidden networks and organizations involved in money laundering. These algorithms can also aid in identifying key players within these groups and help to shed light on the complex structures of money laundering operations. Overall, the use of community detection algorithms can provide valuable insights and assist in the fight against money laundering activities.

5 Conclusion

In conclusion, this study has undertaken a thorough examination of the extant literature on the subject of anti-money laundering, with a particular emphasis on the utilization of machine learning, deep learning, data mining, and big data techniques. Through our review of the relevant literature, we have discerned that while a variety of graph algorithms have been employed, such as graph embedding and centrality algorithms, there appears to be an absence of community detection algorithms that could aid in the identification of organizations and groups engaged in money laundering. This deficit in the literature suggests a promising avenue for future research, wherein the use of community detection algorithms in anti-money laundering systems may be investigated in order to enhance their ability to detect illicit activity. Further-more, our review has also revealed that unsupervised learning methods may prove to be more efficient in the detection of money laundering due to the lack of labeled datasets and the imbalanced nature of such data when it is present. As such, unsupervised methods may represent a viable option for anti-money laundering systems.

References

1. Kute, D.V., et al.: Deep learning and explainable artificial intelligence techniques applied for detecting money laundering—a critical review. *IEEE Access* (2021)
2. Salehi, A., Ghazanfari, M., Fathian, M.: Data mining techniques for anti money laundering. *Int. J. Appl. Eng. Res.* **12**(20), 10084–10094 (2017)
3. Alkhalili, M., Qutqut, M.H., Almasalha, F.: Investigation of applying machine learning for watch-list filtering in anti-money laundering. *IEEE Access* **9**, 18481–18496 (2021)
4. Ketenci, U.G., et al.: A time-frequency based suspicious activity detection for anti-money laundering. *IEEE Access* **9**, 59957–59967 (2021)
5. Martínez-Sánchez, J.F., Cruz-García, S., Venegas-Martínez, F.: Money laundering control in Mexico: a risk management approach through regression trees (data mining). *J. Money Laundering Control* (2020)
6. Jullum, M., Løland, A., Huseby, R.B., Ånonsen, G., Lorentzen, J.: Detecting money laundering transactions with machine learning. *J. Money Laundering Control* (2020)
7. Zhang, Y., Trubey, P.: Machine learning and sampling scheme: an empirical study of money laundering detection. *Comput. Econ.* **54**(3), 1043–1063 (2019)
8. Weber, M., et al.: Anti-money laundering in bitcoin: experimenting with graph convolutional networks for financial forensics. *arXiv preprint [arXiv:1908.02591](https://arxiv.org/abs/1908.02591)* (2019)
9. Colladon, A.F., Remondi, E.: Using social network analysis to prevent money laundering. *Expert Syst. Appl.* **67**, 49–58 (2017)
10. Shokry, A.E.M., Rizka, M.A., Labib, N.M.: Counter terrorism finance by detecting money laundering hidden networks using unsupervised machine learning algorithm. In: *International Conferences ICT, Society, and Human Beings* (2020)
11. Li, X., et al.: Flowscope: spotting money laundering based on graphs. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 4731–4738 (2020)
12. Paula, E.L., et al.: Deep learning anomaly detection as support fraud investigation in Brazilian exports and anti-money laundering. In: *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE (2016)
13. Segovia-Vargas, M.-J.: Money laundering and terrorism financing detection using neural networks and an abnormality indicator. *Expert Syst. Appl.* **169**, 114470 (2021)

14. Ruiz, E.P., Angelis, J.: Combating money laundering with machine learning—applicability of supervised-learning algorithms at cryptocurrency exchanges. *J. Money Laundering Control* (2021)
15. Harris, D.A., et al.: Using real-world transaction data to identify money laundering: Leveraging traditional regression and machine learning techniques. *STEM Fellowship J.* **7**(1), 1–11 (2021)
16. Singh, K., Best, P.: Anti-money laundering: using data visualization to identify suspicious activity. *Int. J. Account. Inf. Syst.* **34**, 100418 (2019)
17. Jensen, R.I.T., Iosifidis, A.: Qualifying and raising anti-money laundering alarms with deep learning. *Expert Syst. Appl.* **214**, 119037 (2023)
18. Lin, C.-Y., Liao, H.-K., Tsai, F.-C.: A systematic review of detecting illicit bitcoin transactions. *Procedia Comput. Sci.* **207**, 3217–3225 (2022)
19. Tertychnyi, P., et al.: Time-aware and interpretable predictive monitoring system for anti-money laundering. *Mach. Learn. Appl.* **8**, 100306 (2022)
20. Han, J., et al.: Artificial intelligence for anti-money laundering: a review and extension. *Digit. Finance* **2**(3), 211–239 (2020)
21. Alarab, I., Prakoonwit, S., Nacer, M.I.: Competence of graph convolutional networks for anti-money laundering in bitcoin blockchain. In: *Proceedings of the 2020 5th International Conference on Machine Learning Technologies* (2020)
22. Hu, Y., et al.: Characterizing and detecting money laundering activities on the bitcoin network. *arXiv preprint [arXiv:1912.12060](https://arxiv.org/abs/1912.12060)*(2019)
23. Lorenz, J., et al.: Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity. In: *Proceedings of the First ACM International Conference on AI in Finance* (2020)



The Reality of Artificial Intelligence Skills Among Eighth-Grade Students in Public Schools

Shatha Sakher¹ , Areeg Al Fouri² , Shatha Al Fouri³,
and Muhammad Turki Alshurideh⁴

¹ Al-Balqa Applied University, As-Salt, Jordan
Shatha.sakher@bau.edu.jo

² Head of the Scientific Research Division, Faculty of Medicine, Al-Balqa Applied University,
As-Salt, Jordan

Areeg.alfouri@bau.edu.jo

³ Jordan University Hospital, Amman, Jordan

⁴ Department of Marketing, School of Business, The University of Jordan, Amman 11942,
Jordan
m.alshurideh@ju.edu.jo

Abstract. This study tends to search the reality of artificial intelligence skills among eighth-grade students in public schools in the city of Al-salt – Jordan, and the descriptive analytical approach was used, where the study tool was developed to measure the skills of artificial intelligence by the researchers, The original community of the study consisted of all students (male and female) in public schools in the city of Al-salt for the first semester of the academic year 2022–2023, and their number is (788) male and female students distributed over (64) schools by using the simple random method. (349) male and female sample students were used. The results indicated that the level of artificial intelligence skills was a moderate degree, Where the dimension of basic skills came in the first place and then advanced skills, and later the dimension of intermediate skills and finally the special skills. Based on the results found, the study recommended that the results of this study should be used to improve students' digital learning for students in the various curricula.

Keywords: Artificial intelligence skills · Public schools in the city of Al-Salt - Jordan

1 Introduction

Educational institutions have witnessed remarkable scientific and technological developments, especially with the respect of the challenges of the Covid-19 pandemic (Ahmad et al. 2021; Leo et al. 2021), which resulted in new transformations in the educational process through the transition to distance learning (Amarneh et al. 2021; Alshurideh et al. 2023), and relying on educational virtual platform (Akour et al. 2021; Al Kurdi et al. 2021a; Alshurideh et al. 2021a). The emergence of special programs and applications that are characterized by diversity and continuous innovation in order to ensure

the progress of the educational process; All of this has enhanced the electronic communication processes in the educational process in all its aspects (Alshurideh et al. 2021b; Sultan et al. 2021).

These processes and transformations witnessed by educational institutions all over the world (Alshurideh et al. 2019a, 2019b), and in Jordan, in particular, it falls under the name of artificial intelligence, Which Muhammad (2021) defines it: automating activities related to human thinking, such as decision-making, and solving problems facing education.

Artificial intelligence is based on the ability to represent knowledge, where special structuring is used to describe knowledge, which includes information and the relationships between this information, and the rules that link these relationships within multiple groups that form the knowledge base in its final form (Nuseir et al. 2021; Alsawan and Alshurideh 2023). Hence, the ability of this knowledge base to improve and develop performance through feedback on errors and weaknesses and neglect some excess information (Abbas 2020; Alhashmi et al. 2020).

According to Cantú-Ortiz et al. (2020), AlShamsi et al. (2021) and Yousuf et al. (2021), artificial intelligence has a number of characteristics such as the ability to perceive and think, acquire knowledge and apply it, learn and understand from previous experiences, use previous experiences and employ them in new situations. The use of trial and error to explore different issues, the ability to respond quickly to new situations and conditions, the ability to deal with difficult and complex cases, the ability to deal with ambiguous situations in the absence of information, the ability to imagine, create, understand and perceive visible matters, and the possibility of providing data to support administrative decisions.

This is what prompted educational institutions to take advantage of new technologies and learn how to manage them optimally (Al Kurdi et al. 2020b; Alshurideh et al. 2021a, 2021b), where this requires the processes of adaptation and integration and bringing about the necessary changes for that (Alshurideh and Kurdi 2023), such as providing students with the necessary skills for artificial intelligence and paying attention to spreading the culture of creativity in various ways (Fabian and Galindo 2021).

2 Research Problem

In the midst of the rapid qualitative development in information technologies; Artificial intelligence systems appeared in the educational field. This has created an important qualitative leap in the behavior of the educational process, and in the way of dealing with applications, platforms, and programs of artificial intelligence, which are smart systems based on knowledge bases, not only in order to represent and store the acquired knowledge, but also with the aim of supporting the knowledge production process (Muhammad 2021). And through the observations of researchers on the skills of artificial intelligence, especially what was associated with the challenges of education in light of the COVID-19 pandemic, and the importance of having artificial intelligence skills for students. In particular that the criteria for the design and evaluation of educational activities are based on the integration modern technologies into education, in order to achieve the completion of the learning activity or part of it, and the importance of this in generating

new ideas, and access to a product or service based on supporting and finding a solution to the problems facing students in the real world (Al-Hamad et al. 2021; Al Shraah et al. 2022; Alshurideh et al. 2022; Nazir et al. 2022).

In this regard, the present study attempts to identify the level of artificial intelligence skills among eighth-grade students in public schools in the city of Al-Salt – Jordan?

Objectives of the Study: This study aimed to identify the level of artificial intelligence skills among eighth-grade students in public schools in the city of Al-Salt - Jordan.

Significance of the Study: This study is importance from two aspects:

- **First: The theoretical aspect:** It is hoped that this study will provide a theoretical framework and previous studies that can be referred to by researchers interested in many of their functional fields about artificial intelligence skills, which will facilitate reference and reliance on it, and it is one of the rare studies in the Arab region as far as the researchers are aware.
- **Second: the practical aspect:** It is hoped that this study will benefit from the provision of highly efficient tools. Acceptable psychometric characteristics related to all skills, facilitate the way for thinkers and researchers in the field of educational psychology and psychological and educational counseling to access them, in light of this, educators, psychological counselors, and officials in the health, psychological, and security fields can benefit from the results of this study to formulate and prepare strategies, programs, and plans to enhance intelligence skills.

3 Conceptual and Procedural Terms

Artificial Intelligence Skills: Malika (2021) defined them as the skills that provide the student with the optimal benefit from the cognitive processors provided by programs, applications, and electronic platforms that are characterized by accuracy, quality, and speed. **While the researchers defined it procedurally:** the degree obtained by the members of the present study sample on the scale of artificial intelligence skills developed by the researchers.

(a) Limitations and determinants of the study

- **Human limits:** This study was limited to public school students in the city of Salt - Jordan in Jordan for all grades, males, and females.
- **Time limits:** This study was applied in the first semester of the academic year 2022–2023.
- **Study limitations:** The present study was limited to applying the study scale (artificial intelligence skills), It is represented by its psychometric properties, through honesty and stability, so the results of this study are determined by the validity of these tools.

(b) Artificial intelligence (AI) skills

Artificial intelligence (AI) represents one of the most important fields of information systems that specializes in developing smart technologies in order to perform tasks or apply solutions through the computer so that the computer has an intelligent behavior to solve problems (Abbas 2020). Abbas (2020) classified the digital skills that a student must have to use artificial intelligence solutions and benefit from its applications, as follows:

- **Basic skills:** These are the skills associated with performing the main tasks and include basic skills for using equipment, such as using the keyboard, touch screen technology, and software, such as managing files on the computer, privacy settings on the phone, and basic online operations, such as operations Search, use e-mail.
- **Intermediate skills:** These are skills that enable students to use artificial intelligence techniques in a more useful and beneficial manner, and include skills related to technology transfer and content creation, such as content publishing, digital design, and digital marketing. These skills are general, meaning that mastering them allows individuals to undertake broader tasks and produce more.
- **Advanced skills:** These are the skills required by specialists to complete their work, such as computer programming, network management, cybersecurity, the Internet of things, and mobile application development. Advanced skills are usually acquired through advanced formal education.
- **Special skills:** These are the skills associated with digital entrepreneurship, which combine traditional entrepreneurship with new digital technology. Which innovates new business models, to deal with customers and stakeholders.

Computer scientist John McCarthy was the first to coin the concept “artificial intelligence” in 1956, and he point out that means the science and engineering of making smart machines, as it was defined as seen by both Kaplan and Haenlein (2019), as the ability of a specific system to analyze any external data. External, deriving new knowledge rules from them and adapting these rules to be used in achieving new goals and tasks.

It is defined as the “specific behavior and characteristics of computer programs, to simulate human mental abilities and patterns of work, such as the ability to learn, deduce and react to new situations” (Mahmoud 2020).

This study believes that artificial intelligence is the electronic method that is based on the practice of human intelligence through computers, so it becomes an effective and highly efficient tool in dealing with educational programs and problems facing students in the education process.

4 Applications of AI in Education

The applications of AI in education can be summarized as mentioned by Al-Farrani and Al-Hujaili (2020) as follows:

- 1- Smart content: where institutions and different digital platforms are interested in creating smart content, by transforming used traditional educational books into new

smart books integrated with audio and video, which may include summaries of study materials and tests, or may include only specific text summaries for each chapter, which are then archived into a digital collection and made available on the dedicated website, in addition to the possibility of self-evaluation.

- 2- These are systems that include educational programs that track student's work and guide them whenever required, by collecting information on the performance of each individual student, and highlighting the strengths and weaknesses of each learner, and providing him with the necessary support in a timely manner.
- 3- Virtual Reality (VR) and Augmented Reality (AR) technology: Where virtual reality (VR) technology works into a computer representation that creates a perception of the world that appears to our senses similar to the real world, which through information and experiences can be transferred to individuals in an attractive and more interactive way, the user feels integrated into the scene, as 3D computer graphics are transformed into virtual environments that can be viewed through multiple browsers, as for augmented reality (AR) technology: It is a type of virtual reality that aimed at replicating the real environment in the computer, It blends the real scene with the virtual scene created by the computer, improving the sensory perception processes of the real world.

5 The Importance of Artificial Intelligence in the Educational Process

AI provided many advantages for developing the educational process especially during the time of the COVID-19 pandemic were mentioned by Mahmoud (2020) as follows:

- The capability to use its tools easily and at different times and places: Artificial intelligence tools are characterized by being suitable in terms of size, such as personal computers laptops and smartphones, which allowed flexibility and ease of use.
- Diversification of applications: Digital applications were characterized by the diversity of fields, activities, specialization, and scientific content, such as mathematics, games, or education..., and the applications varied in terms of targeting the groups to whom they were directed according to the age stages, or according to the educational groups, or others.
- Support for various forms of digital content: that is, the support of these applications for multimedia components, such as sound, video, image and movement, colors and animation text, where these components participate in supporting the presentation and display of content in an interactive manner aimed at attracting the attention of individuals, changing their tendencies and convictions, such as virtual reality technology, augmented reality, holograms, and others.
- Communication Methods: that is, by connecting different devices with each other by easily high-speeding various technologies, starting from wires and even Wi-Fi and Bluetooth technology.
- Simulation of educational environments: Digital technology has been characterized by its ability to build virtual learning environments that are very similar to reality, for example, educational applications combine: the teacher and the student, and the

curriculum within the virtual classroom. in an easy and fast way and in any place, and this contributed to reducing the material cost.

- Providing computing services and storage services: that is, providing technology that is based on transferring and storing data, information, commands, and settings for the user so that information technical programs are transformed from products into services, such as: Google Drive applications, Google Docs, and others.
- Application integration: that is, the ability to use a variety of applications regardless of the type of device, whether smartphones, tablets, or computers, and regardless of their specifications or the operating systems on which they work (Windows – Android – IOS and Mac).

6 Literature Review

Ren et al. (2022) aimed to explore whether family cultural capital may have contributed to the digital inequality of adolescents in relation to both digital skills and uses of digital media. It can illustrate the relationship between the social assets and digital diversity of young people. The cultural capital has been activated as the cultural resource of the family. Cultural practices and parenting activities were related to the media (i.e. active and restricted mediation). The needed data was collected from 1119 middle school students in China were tested. The results presented that cultural practices, cultural resources and active parental mediation were important elements of adolescents' general creative skills and digital skills to use the Internet in education processes. While the use of the Internet in leisure time isn't explained by the cultural capital of the family. The results also denoted a relatively complex pattern of relationships between constrained parental mediation and different dimensions of digital inequality. Track analysis further revealed that cultural resources, cultural practices, and active mediation were key mechanisms for family SES influences on adolescent digital practices. The role of the family's cultural capital in adolescent's digital practices was discussed in the context of media literacy.

A Fabian and Galindo (2021) aimed at analyzing digital skills among a sample of secondary school students based on gender and grade, Using a quantitative approach with a comprehensive descriptive design. The sample included 665 students from public schools in the central province of Peru.

Information was collected through a questionnaire we developed based on recommendations from the National Institute of Educational Technologies and Teacher Training (2017) on digital skills. The survey was grouped into a set of domains which are communication and collaboration, information literacy, information, digital content creation, problem-solving and security. Based on the results, students' digital skill levels had an expected level of achievement with the following outcomes: knowledge and informatics (70.1%), security (61.8%), (61.8%), digital content creation (48.4%), communication and collaboration (47.4%). Digital skills problem-solving processes predominate at 54.3%. It can be concluded that levels of digital skills were prevalent among more than 50% of the secondary schools as expected. It is recommended that teachers integrate these skills to strengthen the learning process and that educational and political authorities implement the technology infrastructure.

Al-Farani and Al-Hujaili (2020) aimed to identify the main element that affecting the teacher's acceptance of the use of AI in education in light of the unified theory of

acceptance and use of technology (UTAUT) to achieve this, the study used the descriptive approach, the theory scale was applied to a sample consisting of (446) male and female teachers in Yanbu Governorate. The results of the study indicated that teachers have a high degree of acceptance of the use of artificial intelligence in education, and that each of (expected performance - expected effort - social impact - available facilities) positively affects the intention to use artificial intelligence in education, and that the most influential factor on the intention of teachers to use artificial intelligence in education is the expected performance followed by the expected effort followed by the social impact followed by the available facilities, The results of the study also indicated that there are statistically significant differences between the sample responses about determining the intention to use artificial intelligence in education due to the gender variable. And These differences were in favor of females. And that there were no statistically significant differences between the sample's responses about determining the intention to use artificial intelligence in education, is attributed to the variable (age, years of experience, and field of educational specialization). In light of these results, the study recommended expanding the use of artificial intelligence applications in education in light of the acceptance of both teachers and learners, and the adoption of the Unified Theory of Acceptance and Use of Technology (UTAUT) to make decisions employing different educational technologies, developing the infrastructure and providing the necessary resources to employ artificial intelligence applications in education.

Mahmoud (2020) aimed at identifying the applications and skills of artificial intelligence that can be used in developing the educational process in light of the challenges of the COVID-19 pandemic, and the study adopted the descriptive approach, For the purposes of the study, a questionnaire was designed. It was presented on behalf of some of those responsible for the educational process in university and pre-university education, and their number reached (107), The study concluded that there are several challenges and problems related to the following aspects: (the educational process - educational administration – teacher-learner - parents – learner's assessment) In light of the COVID-19 crisis, including the limited readiness of teachers and the digital infrastructure in the educational environment, Lack of interest in training teachers and learners to use modern technological techniques. And to fully rely on paper books in the educational process. It concluded that it is possible through the employment of some applications of artificial intelligence in the educational process, such as smart education systems, smart content, virtual reality technology, augmented reality, and others, from facing some of these challenges and problems. And the Study recommended: The need to adopt some applications of artificial intelligence in educational institutions, spreading technological culture and educating educational institutions and the community about it. In addition to, Abbas (2020) study aimed to identify the university student's attitude towards artificial intelligence, and to identify the future orientation among university students, and to identify the relationship between the orientation towards artificial intelligence and the orientation towards the future among the students of Al-Mustansiriya University The study sample consisted of 200 university students, which were drawn using the simple random sampling method. Through the descriptive analytical method, Using the questionnaire as a study tool, the results were as follows: The university students are characterized as having a positive attitude towards artificial intelligence. And that university students

have a positive orientation towards the future, and that there is a statistically significant relationship between the attitude towards artificial intelligence and the orientation towards the future among university students.

7 The Study Approach

This study used the descriptive analytical approach to fulfill the objectives of the study.

7.1 The Community of the Study

Community of the study consisted of all students (males and females) in public schools in the city off Al- Salt for the first semester of the academic year (2022–2023), and they numbered (788) male and female students, distributed over (64) schools, according to the statistics of the annual report of the Ministry of Education. For the year 2019–2020, the study sample was selected according to the sampling table and at level of (0.05) (Al-Najjar et al. 2018), which determines the study sample, which is (349), with a percentage of (10%) of the study population.

7.2 The Study Sample

The study sample was selected from the original study population by simple random method from public schools in the city of Al-Salt - Jordan. The study sample consisted of (349) male and female students. According to the sampling table and rely on the size of the total community the permissible margin of error is (0.05) (Al-Najjar et al. 2018), (349) respondents were reached, and Table 1 shows the distribution of the study sample according to demographic characteristics related to gender.

Table 1. Distribution of the study sample according to the demographic variable (gender).

Variable	Frequency	Percentage
Gender		
Male	169	48.4
Female	180	51.6
Total	349	100.0

7.3 The Study Tools

The scale of artificial intelligence skills was used, and the following are the psychometric characteristics of the scale:

- a. **The AI tool (scale)** was built with reference to theoretical framework and literature review It may be the measure of artificial intelligence skills, which consists of (12) items, distributed over (4) dimensions.

- b. **The construct validity (CV):** To ensure the validity of the study tool and to measure the test variables, several procedures were taken, Apparent validity: in order to ensure the content validity of the study tool, it was presented in its initial form to (12) arbitrators with expertise and specialization from faculty members in government and private Jordanian universities, In order to determine the appropriateness of the tool to fulfill the goals of the study, This is to ensure its suitability to measure what it should measure and the suitability and affiliation of the paragraphs to measure the variables of the study. And suggesting any appropriate paragraphs that could enrich the study, Ensure the clarity and linguistic correctness of the existing paragraphs and express their opinion on the form, coordination and method of measurement, their comments on the tool have been taken, The paragraphs were modified, deleted and proofread until they reached their final form.
- c. **The stability of the study tool:** The stability of the study tool was confirmed based on the results of Cronbach's Alpha test values, it is a test that measures the internal consistency of the respondent's responses that were distributed to them, or it is a statistical test through which the internal consistency of the answers of the persons to whom the study tool was distributed is clarified. To ensure the stability of the study tool, the correlation coefficients were calculated for each of the paragraphs belonging to each dimension of the independent factor, in addition to the dependent paragraphs of each dimension of the dependent factor, The Cronbach alpha value was also calculated for the total paragraphs for each of the strategic planning, and creative behavior. Table 2 shows the results of the stability coefficients for the items of the study tool using the Cronbach alpha test.

Table 2. The stability coefficients of study instrument paragraphs employing Cronbach Alpha test

Variable	The values of the validity and reliability of the questionnaire - Alpha Cronbach test values
Basic skills	0.755
Intermediate skills	0.766
Advanced skills	0.725
Special skills	0.753
Artificial intelligence skills	0.794

- d. **Scale correction key:** It has been taken into account that the 5-scale used in the study is graded according to the rules and characteristics of the scales as follows:

Answer alternatives				
Strong agree (SA)	Agree (A)	Neutral (N)	Disagree (D)	Strong disagree (SD)
1	2	3	4	5

Based on the above, the values of the arithmetic averages that the study reached were dealt with as follows:

- The low level is $1.00 + 1.33 = 2.33$
- The low level is $1.00 + 1.33 = 2.33$
- Intermediate level is $3.67 = 1.33 + 2.34$
- The high level is from 5.00–3.68

8 The Answer to the Study Questions

The first question: What is the level of artificial intelligence skills among public school students in the city of Al-salt - Jordan?

To answer this question, the arithmetic averages and deviations for the field of artificial intelligence skills were extracted, and the table below shows that?

Table 3. Arithmetic averages and standard deviations of the role of cultural heritage in preserving Arab cultural identity, in descending order by Arithmetic averages.

No	The field	Arithmetic average (AA)	Standard deviation (SD)	Rank (R)	Level (L)
1	Basic skills	3.80	1.01	1	High
2	Intermediate skills	2.87	1.208	3	Average
3	Advanced skills	2.94	0.897	2	Average
4	Special skills	2.17	0.856	4	Weak
	Total marks	2.95	1.015	Average	

Table 3 shows that the arithmetic averages ranged between (3.80–2.17), where the basic skills came in the first rank with the highest arithmetic average (3.80), while special skills came in last place with an arithmetic average of (2.17), and the arithmetic average of the tool as a whole was (2.95). The arithmetic averages and standard deviations of the estimates of the study sample members were calculated in the paragraphs of each field separately, as follows:

- 1- **Basic skills:** The arithmetic average and standard deviations of the responses of the study sample were extracted to identify the level of basic skills among public school students in the city of Salt - Jordan, and Table 4 illustrates this.

Table 4. The arithmetic average and standard deviations of the study sample's responses to the "Basic Skills" paragraphs are arranged in descending order.

No.	Paragraph	Arithmetic average	Standard deviation	Arrangement	Level
1	I have the ability to manage files on the computer	3.87	0.90	1	High
2	I can modify the privacy settings on the phone	3.79	1.13	2	High
3	Have the skills of online search and e-mail	3.76	1.02	3	High
	The general arithmetic average	3.80	1.01	High	

Based on Table 4, the arithmetic averages for the (basic skills) was ranged between (3.87 and 3.46), whereas the basic skills of public school students in the city of Al-salt - Jordan had a total arithmetic average of (3.80), which is of the high-level, Paragraph No. (1) obtained the highest arithmetic average, reaching (3.87), with a standard deviation of (0.90), which is from the high level, The paragraph stated (I have the ability to manage files on the computer), and in the second place came Paragraph No. (2).) with an arithmetic average (3.79) and a standard deviation of (1.13), which is from the high level, as the paragraph states (I can modify the privacy settings on the phone), and in the last place came paragraph No. (3) with an arithmetic average (3.46) and a standard deviation (1.21), which is from the average level, as the paragraph stipulated(I have the skills of online search and e-mail).

- 2- **Intermediate skills:** Arithmetic average and standard deviations of the responses of the study sample were extracted to identify the level of intermediate skills among public school students in the city of Al-salt - Jordan, and Table 5 illustrates this.

Based on Table 5, the Arithmetic Averages (AA) for the intermediate skills were ranged between (2.97 and 2.89), where the average skills were acquired by students of public schools in the city of Al-salt - Jordan, which is of the average level, and the paragraph stipulates (master the digital design of images, videos, and Photoshop processes), In the second place, came Paragraph No. (6) with an arithmetic average of (2.89) and a standard deviation of (1.186). It is from the average level, as the paragraph stipulates (I can publish appropriate electronic content on electronic applications and platforms). And in the last place came paragraph No. (5) with an arithmetic average (2.77) and a standard deviation (1.285), which is from the average level, as the paragraph stated (I have knowledge of digital marketing techniques and the mechanisms of their application).

Table 5. The arithmetic means and standard deviations of the study sample's responses to the "intermediate skills" items, arranged in descending order.

No	Paragraph	Arithmetic average (AA)	Standard deviation (SD)	Arrangement (A)	Level (L)
4	Master the digital design of photos, videos, and Photoshop operations	2.97	1.153	1	Average
5	I have knowledge of digital marketing techniques and the mechanisms of their application	2.77	1.285	3	Average
6	I can publish appropriate electronic content on electronic applications and platforms	2.89	1.186	2	Average
	The general arithmetic average	2.87	1.208		

3- **The advanced skills:** Arithmetic means and standard deviations of the responses of the study sample were extracted to identify the level of advanced skills among public school students in the city of Salt - Jordan, and Table 6 illustrates this.

As see in Table 6, the Arithmetic Average (AA) for the advanced skills were ranged between (2.35 and 3.66). As the advanced skills of public-school students in the city of Al-salt - Jordan obtained a total arithmetic average of (2.94), which is of the average level. Paragraph No. (8) had the highest arithmetic average, reaching (3.66), with a standard deviation of (0.531), which is from the average level. The paragraph stated (I can manage electronic networks and develop electronic applications), and in the second place came Paragraph No. (7) with an arithmetic mean of (2.83) and a standard deviation of (1.262), which is from the average level, which is from the average level, as the paragraph stipulates (I have the ability to program websites).

and in the last place came paragraph No. (9) with an arithmetic average of (2.35) and a standard deviation of (0.90), which is from the average level, as the paragraph stated (I have the ability to activate cybersecurity operations and the Internet of things).

4- **Special skills:** The arithmetic average and standard deviations of the responses of the study sample were extracted to identify the level of special skills among public school students in the city of AL-Salt - Jordan, and Table 7 illustrates this.

Table 6.

No	Paragraph	Arithmetic Average (AA)	Standard deviation (SD)	Arrangement (A)	Level (L)
7	I have the ability to program websites	2.83	1.262	2	Average
8	I can manage electronic networks and develop electronic applications	3.66	0.531	1	Average
9	I have the ability to activate cyber security operations and the Internet of things	2.35	0.90	3	Average
	The general arithmetic average	2.94	0.897	Average	

Table 7. The arithmetic average and standard deviations of the study sample's responses to the "special skills" items, arranged in descending order.

No	Paragraph	Arithmetic average (AA)	Standard deviation (SD)	Arrangement (A)	Level (L)
10	I obtained a training certificate in the field of digital entrepreneurship	2.01	0.77	3	Weak
11	Participated in competitions for digital entrepreneurship	2.20	0.84	2	Weak
12	I have the ability to generate innovation for new digital business models	2.30	0.96	1	Weak
13	The general arithmetic average	2.17	0.856	Weak	

It is clear from Table 7 that the Arithmetic Average (AA) for (special skills) ranged between (2.30 and 2.01), where the special skills of public-school students in the city

of Al-salt - Jordan obtained a total arithmetic average of (2.17), which is of the weak level. Paragraph No. (12) Obtained the highest arithmetic average, reaching (2.30), with a standard deviation of (0.96), which is from the weak level. The paragraph stated (I have the ability to generate innovation for new digital business models), and in the third place came Paragraph No. (11) with an arithmetic average of (2.20) and a standard deviation of (0.84), which is from the weak level, as the paragraph stipulates (I participated in competitions for digital entrepreneurship), and in the third place came paragraph No. (11) with an arithmetic average of (2.01) and a standard deviation of (0.77), which is from the weak level, as the paragraph stated (I obtained a training certificate in the field of digital entrepreneurship).

9 Discussion of Results and Recommendations

Results of the study denoted that the level of AI skills among eighth-grade students in public schools in the city of Al-salt - Jordan came at an average level, where the basic skills dimension came in the first place, the advanced skills dimension in the second place, then the intermediate skills dimension, and finally the special skills dimension. Because there aren't any training courses for students in public schools on cybersecurity operations, digital innovation models, and entrepreneurship training courses. The researchers recommend the following:

- Introducing modern models of artificial intelligence skills in the teacher's guide in schools.
- Participating with the private sector to provide courses for students in public schools on digital business, and to involve them in competitions related to entrepreneurship and digital innovation models.
- Activating partnerships with security authorities and technical agencies specialized in cybersecurity to provide awareness and training services for students in public schools.
- Using the results of this study to improve digital learning for students in the various courses, which will contribute to the transfer of knowledge and keep abreast of digital developments.
- Developing the infrastructure and providing the necessary resources to integrate artificial intelligence applications in schools, and promoting digital resources in public schools in general, by increasing investment in designing educational simulation programs, artificial intelligence, virtual reality technologies, and augmented reality to benefit from them in teaching.

References

- Malika, M.: Artificial intelligence and the future of distance education. *Stud. Dev. Soc.* **6**(3), 131–144 (2021)

- Ahmad, A., Alshurideh, M.T., Al Kurdi, B.H., Salloum, S.A.: Factors impacts organization digital transformation and organization decision making during Covid19 pandemic. In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence*. SSDC, vol. 334, pp. 95–106. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67151-8_6
- Akour, I., Alshurideh, M., Al Kurdi, B., Al Ali, A., Salloum, S.: Using machine learning algorithms to predict people's intention to use mobile learning platforms during the COVID-19 pandemic: machine learning approach. *JMIR Med. Educ.* **7**(1), 1–17 (2021)
- Al-Farani, L., Al-Hujaili, S.: Factors affecting teacher acceptance of the use of artificial intelligence in education in light of the unified theory for acceptance and use of technology (UTAUT). *Arab J. Educ. Psychol. Sci.* **14**, 215–252 (2020)
- Al-Hamad, M., Mbaidin, H., AlHamad, A., Alshurideh, M., Kurdi, B., Al-Hamad, N.: Investigating students' behavioral intention to use mobile learning in higher education in UAE during Coronavirus-19 pandemic. *Int. J. Data Netw. Sc.* **5**(3), 321–330 (2021)
- Alhashmi, S.F.S., Alshurideh, M., Al Kurdi, B., Salloum, S.A.: A Systematic review of the factors affecting the artificial intelligence implementation in the health care sector. In: Hassanien, A.E., Azar, A., Gaber, T., Oliva, D., Tolba, F. (eds.) *AICV 2020. AISC*, vol. 1153, pp. 37–49. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44289-7_4
- Al Kurdi, B., Alshurideh, M., Salloum, S.A.: Investigating a theoretical framework for e-learning technology acceptance. *Int. J. Electr. Comput. Eng. (IJECE)* **10**(6), 6484–6496 (2020a)
- Al Kurdi, B., Alshurideh, M., Salloum, S., Obeidat, Z., Al-dweeri, R.: An empirical investigation into examination of factors influencing university students' behavior towards elearning acceptance using SEM approach. *Int. J. Interact. Mob. Technol.* **14**(2), 19–24 (2020b)
- Al-Najjar, F.J., Al-Najjar, N.J., Al-Zoubi, M.R.: *Methods of Scientific Research: An Applied Perspective*, 4th edn. Dar Al-Hamid for Publication and Distribution, Amman, Jordan (2018)
- Al Shraah, A., Abu-Rumman, A., Alqhaiwi, L., Alshurideh, M.T.: The role of AACSB accreditation in students' leadership motivation and students' citizenship motivation: business education perspective. *J. Appl. Res. High. Educ.* (ahead-of-print) (2022)
- Alsawan, N.M., Alshurideh, M.T.: The application of artificial intelligence in real estate valuation: a systematic review. In: Hassanien, A.E., Snášel, V., Tang, M., Sung, T.W., Chang, K.C. (eds.) *AISI 2022. LNDECT*, vol. 152, pp. 133–149. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-20601-6_11
- AlShamsi, M., Salloum, S.A., Alshurideh, M., Abdallah, S.: Artificial intelligence and blockchain for transparency in governance. In: Hassanien, A., Bhatnagar, R., Darwish, A. (eds.) *Artificial Intelligence for Sustainable Development: Theory, Practice and Future Applications*. SCI, vol. 912, pp. 219–230. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-51920-9_11
- Alshurideh, M., Al Kurdi, B., Salloum, S.A.: Examining the main mobile learning system drivers' effects: a mix empirical examination of both the expectation-confirmation model (ECM) and the technology acceptance model (TAM). In: Hassanien, A., Shaalan, K., Tolba, M. (eds.) *AISI 2019. AISC*, vol. 1058, pp. 406–417. Springer, Cham (2019a). https://doi.org/10.1007/978-3-030-31129-2_37
- Alshurideh, M., Salloum, S.A., Al Kurdi, B., Monem, A.A., Shaalan, K.: Understanding the quality determinants that influence the intention to use the mobile learning platforms: a practical study. *Int. J. Interact. Mob. Technol.* **13**(11), 183–157 (2019b)
- Alshurideh, H., Al Kurdi, B., Alshurideh, M., Alkurdi, S.: Covid-19 pandemic and students' life: the impact on employment opportunities. *Int. J. Theory Organ. Pract. (IJTOP)* **2**(1), 80–98 (2022)
- Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.): *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence*. SSDC, vol. 334. Springer, Cham (2021a). <https://doi.org/10.1007/978-3-030-67151-8>

- Alshurideh, M.T., et al.: Factors affecting the use of smart mobile examination platforms by universities' postgraduate students during the COVID 19 pandemic: an empirical study. *Informatics* **8**(2), 1–21 (2021b)
- Alshurideh, M., Kurdi, B.: Factors affecting social networks acceptance: an extension to the technology acceptance model using PLS-SEM and machine learning approach. *Int. J. Data Netw. Sci.* **7**(1), 489–494 (2023)
- Alshurideh, M., Abuanzeh, A., Kurdi, B., Akour, I., AlHamad, A.: The effect of teaching methods on university students' intention to use online learning: technology acceptance model (TAM) validation and testing. *Int. J. Data Netw. Sci.* **7**(1), 235–250 (2023)
- Alshurideh, M., Al Kurdi, B., Salloum, S.A., Arpaci, I., Al-Emran, M.: Predicting the actual use of m-learning systems: a comparative approach using PLS-SEM and machine learning algorithms. *Interact. Learn. Environ.*, 1–15 (2020b)
- Amarneh, B.M., Alshurideh, M.T., Al Kurdi, B.H., Obeidat, Z.: The impact of COVID-19 on e-learning: advantages and challenges. In: Hassanien, A.E., et al. (eds.) *AICV 2021. AISC*, vol. 1377, pp. 75–89. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_8
- Muhammad, A.: The role of artificial intelligence applications in improving the quality of educational service, a survey of the opinions of a sample of employees at Al-Furat Al-Awsat technical university. *J. Fac. Adm. Econ. Econ. Adm. Financial Stud.* **13**(1), 127–154 (2021)
- Abbas, R.: The trend towards artificial intelligence and its relationship to the future orientation of university students. *J. Arts* **135**, 367–406 (2020)
- Leo, S., Alsharari, N.M., Abbas, J., Alshurideh, M.T.: From offline to online learning: a qualitative study of challenges and opportunities as a response to the COVID-19 pandemic in the UAE higher education context. In: Alshurideh, M., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence. SSDC*, vol. 334, pp. 203–217. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67151-8_12
- Nazir, J., et al.: Perceived factors affecting students academic performance. *Acad. Strateg. Manag. J.* **21**(Spec. Issue 4), 1–15 (2022)
- Nuseir, M.T., Al Kurdi, B.H., Alshurideh, M.T., Alzoubi, H.M.: Gender discrimination at workplace: do artificial intelligence (AI) and machine learning (ML) have opinions about it. In: Hassanien, A.E., et al. (eds.) *AICV 2021. AISC*, vol. 1377, pp. 301–316. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_28
- Mahmoud, A.-R.: Applications of artificial intelligence: an introduction to education development in light of the challenges of the COVID 19 pandemic. *Int. J. Res. Educ. Sci.* **3**(4), 171–224 (2020)
- Cantú-Ortiz, F.J., Galeano Sánchez, N., Garrido, L., Terashima-Marin, H., Brena, R.F.: An artificial intelligence educational strategy for the digital transformation. *Int. J. Interact. Des. Manuf. (IJIDeM)* **14**(4), 1195–1209 (2020). <https://doi.org/10.1007/s12008-020-00702-8>
- Kaplan, A., Haenlein, M.: Siri, Siri in my hand, who's the fairest in the land? On the interpretations, illustrations and implications of artificial intelligence. *Bus. Horiz.* **62**(1), 15–25 (2019)
- Sultan, R.A., Alqallaf, A.K., Alzarooni, S.A., Alrahma, N.H., AlAli, M.A., Alshurideh, M.T.: How students influence faculty satisfaction with online courses and do the age of faculty matter. In: Hassanien, A.E., et al. (eds.) *AICV 2021. AISC*, vol. 1377, pp. 823–837. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_72
- Yousuf, H., Zainal, A.Y., Alshurideh, M., Salloum, S.A.: Artificial intelligence models in power system analysis. In: Hassanien, A., Bhatnagar, R., Darwish, A. (eds.) *Artificial Intelligence for Sustainable Development: Theory, Practice and Future Applications. SCI*, vol. 912, pp. 231–242. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-51920-9_12
- Ren, W., Zhu, X., Yang, J.: The SES-based difference of adolescents' digital skills and usages: an explanation from family cultural capital. *Comput. Educ.* (2022)



A Deep Neural Network Architecture for Extracting Contextual Information

Zakariae Alami Merrouni¹ (✉), Bouchra Frikh¹, and Brahim Ouhbi²

¹ LIASSE Lab, National School of Applied Sciences (ENSA), Sidi Mohamed Ben Abdellah University Fez, Fez, Morocco

zakariae.alamimerrouni@usmba.ac.ma, bfrikh@yahoo.com

² LM2I Lab, National Higher School of Arts and Crafts (ENSAM), Moulay Ismail University, Meknes, Morocco

Abstract. The exponential growth of textual data makes it challenging to retrieve pertinent information. Numerous methods for automating keyphrase extraction have emerged from earlier studies. Keyphrases have been used extensively to analyze, organize, and retrieve text content across various domains. Previous works have yielded numerous viable strategies for automated keyphrase extraction. They rely on domain-specific knowledge and features and select and rank the most relevant keyphrases. In this paper, we propose a deep neural network architecture based on word embedding and a Bidirectional Long Short-Term Memory Recurrent Neural Network “Bi-LTSM”. This architecture can capture the hidden context and the main topics of the document. Experimental analysis of benchmark datasets reveals that our proposed model achieves noteworthy performance compared to baselines and previous approaches for keyphrase extraction.

Keywords: Neural network · Bi-LTSM · Word embedding · Contextual information · Keyphrase extraction

1 Introduction

The exponential growth of textual data and online sources necessitates the use of automatic processes to find pertinent embedded information. Keyphrases can be used to express this information. They describe the main topics of a document [1]. The task of automatically extracting relevant and descriptive keyphrases from a document or web data is known as automatic key extraction (AKE). Over the past two decades, a lot of attention has been received to this task [1]. AKE has been successfully used in many Natural Language Processing (NLP) and Information Retrieval (IR) tasks such as document clustering [2], text summarization [3], opinion mining [4], sentiments analysis [5], information extraction [6], and recommender systems [7].

Early AKE systems were based on supervised machine learning algorithms, such as SVM [8], bagged C4.5 [9], Naïve Bayes [10]. The majority of supervised methods have focused on (1) selecting features that measure candidate keyphrases and (2) running learning algorithms to distinguish keyphrases from non-keyphrases [11]. Statistical,

semantic, linguistic, and external features are all used in these supervised algorithms [1]. Unsupervised AKE algorithms have also been developed, in which the AKPE task is viewed as a ranking problem and performed without prior knowledge. The majority of unsupervised keyphrase extraction techniques are classified as statistical, graph-based, or embeddings-based [1, 11]. For instance, in KeyphraseDS system [12], Keyphrases are extracted using a conditional random field (CRF)-based model that takes advantage of a variety of features. In [13], The authors present RVA, a novel unsupervised method for keyphrase extraction that makes use of local word embeddings (in particular GloVe vectors).

Nonetheless, given the wide range of features that can contribute to the definition of a keyphrases, AKE's performance on the most common evaluation sets remains moderate [1], and the NLP community still has a lot of room for improvement. In recent years, Deep Learning techniques have shown impressive results in many NLP tasks, including Sentiment Analysis, Machine Translation, Automatic Summarization, Named Entity Recognition, Question Answering, and so on [1]. Some attempts have recently been made to handle the AKE task using Deep Learning approaches [14–20]. In this paper, we present a novel deep learning-based architecture for extracting relevant information via the AKE. The proposed approach, in particular, is based on a Bidirectional Long Short-Term Memory Recurrent Neural Network (Bi-LSTM) and a word embedding approach, which can capture the semantic and context of a given word. The architecture of our system is adaptable to multiple domains and can be used in a variety of situations. Experiments are carried out on three well-known datasets to evaluate the proposed method. The experimental results show that the proposed architecture outperforms some well-known approaches.

2 Related Work

In previous studies, the automatic keyphrase extraction task was divided into two steps: (1) using heuristics such as POS patterns for words or n-grams to identify a set of proper phrases or words as candidate keyphrases, (2) the candidate phrases are predicted as keyphrases or non-keyphrases using a supervised or unsupervised approach [1]. Statistical, semantic, linguistic, and/or external features are used in supervised approaches to assess the relevance of these keyphrases. Then, algorithms such as SVM [8], bagged C4.5 [9], Naïve Bayes [10] are used to classify them. Unsupervised approaches include metric scoring, graph-based techniques, topic modeling, and embeddings-based techniques. Several supervised and unsupervised approaches are compared and analyzed in a survey by Merrouni et al. [1].

Deep Learning techniques have recently demonstrated impressive results in information extraction tasks. For instance, Wang et al. [21] suggest using word embeddings to assess the relatedness of words in graph-based models. Zhang et al. [15] uses a Recurrent Neural Network (RNN) based approach to identify keyphrases in Twitter data. Their model addresses the problem of sequence labeling for very short text using two layers, with the joint-layer RNN used to capture the semantic dependencies in the input sequence [15]. Augenstein and Søgaard [22] handle keyphrase extraction as a multi-task learning problem, and use RNNs to classify keyphrase boundaries with various parameters.

Meng et al. [14] focused on keyphrase generation (rather than extraction) by compressing the original text into a hidden layer with an encoder and predicting a keyphrase with a decoder, where the sequence of words in a document is used to generate a sequence of keyphrases. However, this method suffered from the prediction of various semantical phrases. To generate the keyphrases sequences, an original Encoder-Decoder RNN was proposed by Cho et al. [26]. A variant of the model that outperforms the Encoder-Decoder RNN includes a copying mechanism to identify keyphrases that appear infrequently in the text. Chen et al. [23] proposed CorrRNN to generate keyphrases that capture correlation between multiple keyphrases. This method removes semantically duplicated keyphrases. Nevertheless, this method necessitates a massive amount of labeled data for training. Wang et al. [24] proposed a novel Topic-based Adversarial Neural Network (TANN) method for utilizing unlabeled data in the target domain as well as data in the resource-rich source domain. This method reduces the amount of training data required and improves keyphrase extraction performance on unlabeled or inadequately labeled target domains. Basaldella et al. [25] exploit the preceding and following context of a given phrase to predict keyphrases, to this end, a (BiLSTM) RNN network is used. Santosh et al. [27] proposed (DAKE), a document-level attention model for extracting keyphrases from scientific papers. BiLSTM and CRF are combined in this model. Wu et al. [28] propose a method based on the keyphrases that relies on the use of keywords mentioned in the text to construct the missing ones using a mask-predict method.

3 Proposed Architecture

In this section, we present our architecture for extracting relevant information using a deep neural AKE approach.

To extract keyphrases, we took the following steps (Fig. 1): (1) Sentence splitting, (2) text pre-processing, and (3) word embedding, as well as (4) neural keyphrase extraction. Each stage is described as follows:

3.1 Preprocessing

Sentence Splitting

In the first step, we divided the input document into sentences. Sentence splitting is a common preliminary step on texts before further processing, which is known as sentence boundary detection or sentence segmentation. It is the process of breaking down the input text into sentences. We employ the Stanford Parser [39] for this aim.

Text Preprocessing

In the second step, we pre-process the sentences using the following stages:

- Tokenizing: we tokenize each sentence by removing all marks, punctuation, brackets, numbers, and special characters, as well as lower-casing all words. For example, given the sentence: (a-deep learInstA~ance;;;,; ing;is;mo-del^*is an = rnn), the result will be: (a deep learning model is an rnn).

- **Normalizing:** Normalization is the process of converting a token to its base form. During the normalization process, the inflectional form of a word is removed so that the base form can be obtained. Normalization is useful for reducing the number of unique tokens and removing variations in a text. In addition to cleaning up the text by removing superfluous information. We use the Stanford CoreNLP library for Normalizing our sentences [39].

Word Embedding

Because our model's input layer is a vector representation of each word in the input document, we want words with similar context and semantics to have similar representations. In a predefined vector representation, the words are represented as real-valued vectors. Each input word is assigned to a continuous vector, and the vector values are learned in a manner similar to the proposed neural network architecture.

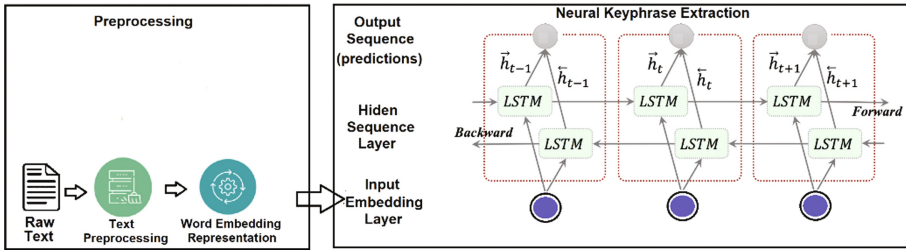


Fig. 1. Overview of the proposed system's Architecture.

For efficiently learning word vectors, we used the Global Vectors for Word Representation “GloVe” algorithm [30], which is an extension to the word2vec method. GloVe is a method for combining global statistics from matrix factorization techniques such as LSA with local context-based learning in word2vec. Rather than defining local context through a window, GloVe creates an explicit word-context or word co-occurrence matrix using statistics from the entire text corpus. To improve efficiency, the embedded words are trained on large unlabeled data sets to represent each word's context and semantics [29]. This extensive training has resulted in a significant improvement in these embeddings [29]. However, existing datasets for AKE are quite small, so we use Stanford's GloVe Embeddings, which are trained on either Wikipedia 2014 + Gigaword 5 (approximately 6 billion tokens) or Common Crawl (approximately 6 billion tokens) (about 840 billion tokens) [26]. GloVe's training goal is to obtain word vectors whose dot product equals the logarithm of the probability of the words occurring together. Because the logarithm of a ratio equals the difference of logarithms, this objective connects ratios of co-occurrence probability with vector differences in the word vector space. This information is also recorded as vector disparities since these ratios can encode some level of significance. As a result, the produced word vectors performed well on word analogy tasks such as those investigated by the word2vec algorithm.

3.2 Neural Keyphrase Extraction Architecture

Bidirectional long-short term memory (Bi-LSTM) is the process of constructing any neural network with sequence information that can be read backwards (future to past) or forwards (past to future). Our input of word embeddings flows in two directions in bidirectional, distinguishing a Bi-LSTM from a regular LSTM. This input flow in both directions preserves future and past information, and it can cope with variable sentence lengths and assess word features and their context successfully. Our Bi-LSTM is linked to a fully connected hidden layer, which is linked to a softmax-POS output layer with POS embedding input and three neurons per word. Using softmax-POS, a probability distribution for the possible tags is generated for each token in the sentence. The tag with the highest likelihood is chosen. The three neurons are mapped to three possible output classes: N-KP, B-KP, and IN-KP, which respectively mark non-keyphrase tokens, the first token of a keyphrase, and the remaining tokens of a keyphrase.

Bi-LSTM Model

Long Short-Term Memory Networks (LSTMs) are a subset of Recurrent Neural Networks (RNNs) that are designed to overcome RNNs' gradient vanishing problem. LSTMs, in particular, have additional memory cells that store memory from long distance terms [32]. Since LSTMs can retain information from previous sequence inputs into the current input state, they are a natural choice for applications involving temporal and sequence data, such as speech recognition, language modeling, and translation [25, 26, 28].

An LSTM has three layers: an input layer, a hidden layer, and an output layer. The sequential information is captured using a Bi-LSTM. At time t , an LSTM unit (see Fig. 2) consists of sub-unit inputs (i_t), outputs (o_t), forget gates (f_t), and memory cells (c_t).

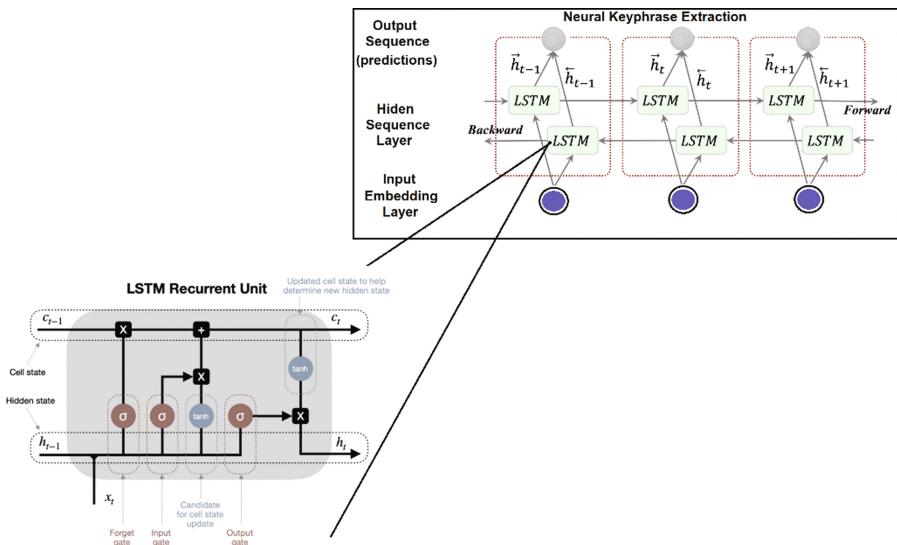


Fig. 2. Long Short-Term Memory (LSTM) Neural Networks cell general architecture

Our LSTMs, in particular, compute their state sequence ($h_1, h_2, \dots, h_n$) at time-step t by following these equations given input vectors ($x_1, x_2, \dots, x_n$) which are the word embeddings representation:

- $\mathbf{i}_t = \sigma(\mathbf{U}^{(i)}\mathbf{x}_t + \mathbf{W}^{(i)}\mathbf{h}_{t-1} + \mathbf{W}^{(i)}\mathbf{c}_{t-1} + \mathbf{b}^{(i)})$
- $\mathbf{f}_t = \sigma(\mathbf{U}^{(f)}\mathbf{x}_t + \mathbf{W}^{(f)}\mathbf{h}_{t-1} + \mathbf{W}^{(f)}\mathbf{c}_{t-1} + \mathbf{b}^{(f)})$
- $\mathbf{o}_t = \sigma(\mathbf{U}^{(o)}\mathbf{x}_t + \mathbf{W}^{(o)}\mathbf{h}_{t-1} + \mathbf{W}^{(o)}\mathbf{c}_t + \mathbf{b}^{(o)})$
- $\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}^{(c)}\mathbf{x}_t + \mathbf{W}^{(c)}\mathbf{h}_{t-1} + \mathbf{b}^{(c)})$
- $\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t)$

where, σ is the element-wise logistic sigmoid function and $\odot$ is the Hadamard product; i, f, o , and c are the input gate, forget gate, output gate and cell activation vectors, $\mathbf{h}_t$ is the hidden state that stores sequential information up to time t . $\mathbf{x}_t$ is the input vector at time t , i.e., considered as the word embedding in our model, and $\mathbf{W}$ and $\mathbf{b}$ are learned biases.

We will use a Bi-LSTM model (with a forward hidden layer and a backward hidden layer) to exploit not only a word's previous context, but also its future context. The structure of a Bi-LSTM network is depicted in Fig. 2. This model's architecture is made up of two distinct hidden layers. It computes the forward hidden sequence $\vec{h}_t$ first; then the backward hidden sequence $\overleftarrow{h}_t$, and finally the backward and forward hidden sequences are combined to generate the output y_t .

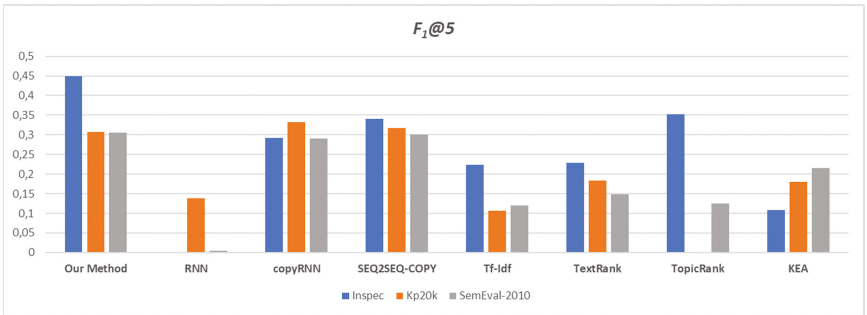
4 Experimental Results and Discussion

We run a series of experiments on three well-known keyphrase extraction datasets to assess the performance of our proposed model. These Datasets are based on multiple domains. The first dataset INSPEC [33] is made up of 2000 abstract papers in English extracted from journal papers in the fields of Computer and Control, Information Technology. It is made up of 1000 documents for training, 500 for validation, and 500 for testing. The second dataset kp20k [14] crawled paper metadata from multiple online digital libraries, including ACM, WebofScience, and Wiley. The kp20k contains metadata for 567,830 papers that are clearly divided into train, validation, and test sets provided by the authors, as shown below: 527,830 were used for model training, 20,000 for parameter adjustment, and 20,000 for model evaluation. The third dataset is SemEval-2010 [38], which contains 284 full-length ACM articles (144 for training, 40 for validation, and 100 for testing) with both abstractive and extractive keyphrases. After experimenting with several network configuration, we found that 200 neurons for the Bi-LSTM layer, 200 neurons for the hidden dense layer, and 0.25 for the dropout layers in between produced the best results. The learning rate was set to 0.07 and the models were trained for a total of 66 epochs with patience value of 5 and annealing factor of 0.5. One of the main goals of this work is to investigate the effectiveness of contextual embeddings in keyphrase extraction. To assess the influence of word embeddings, we run tests with pre-trained Stanford's GloVe Embeddings in all word embedding sizes (50, 100, 200, and 300). The number of epochs requested to converge in all the four settings is 25,15,19,7 respectively. After experimenting with various settings, we discovered that the best results are obtained with an embedding size of 100 and an increase in precision [31].

Table 1. Comparison of our model with previous approaches on Kp20k, Inspec, and SemEval-2010 datasets

Method	Inspec	Kp20k	SemEval-2010
	$F_1@5$	$F_1@5$	$F_1@5$
Our Method	0.450	0.308	0.305
<i>RNN</i>	0.000	0.138	0.004
<i>copyRNN</i>	0.292	0.332	0.291
<i>SEQ2SEQ-COPY</i>	0.340	0.318	0.301
<i>Tf-Idf</i>	0.223	0.107	0.120
<i>TextRank</i>	0.229	0.183	0.149
<i>TopicRank</i>	0.352	N/A	0.125
<i>KEA</i>	0.109	0.180	0.215

As a result, in Table 1, we compared the best resulted setting of our system (in terms of F-score) to some baseline methods such as: copyRNN [35], a recently proposed deep learning model, is a sequence-to-sequence learning model based on an RNN Encoder-Decoder framework combined with a copying mechanism. SEQ2SEQ-COPY [17], a semi-supervised learning framework based on seq2seq models that leverages unlabeled data for keyphrase generation. TfIdf, TextRank [36], and TopicRank [34] are graph-based keyphrase extraction methods that rely on a topical representation of the document. KEA is a supervised model [37]. The results show that our proposed model performs well in terms of F1-score. It also demonstrated that unsupervised models (Tf-idf, TextRank) perform well across different datasets (see Fig. 3).

**Fig. 3.** The F1 scores of the top 5 keyphrases on the on the KP20K, INSPEC, and SEMEVAL-2010 DATASETS

Due to the high time complexity of the Kp20k dataset, the TopicRank fails to return any results, but it performs well on the Inspec dataset. Furthermore, our method outperforms the compared deep neural methods, such as copyRNN and SEQ2SEQ-COPY on the Inspec and SemEval-2010 datasets. In the Inspec dataset, for example, our

method achieves an F1-score of 45.0%, compared to 29.2% for copyRNN and 34.0% for SEQ2SEQ-COPY. However, SEQ2SEQ-COPY (F1-score of 31.8%) and copyRNN (F1-score of 32.2%) outperform our method (F1-score of 30.8%) by a slight margin on the Kp20k dataset. This is because both methods use labeled data as well as large-scale unlabeled samples for learning and generate new keyphrases that did not appear in the text. It is regarded as a keyphrase generation task rather than an extraction task. As for the RNN model with the attention mechanism did not perform as well as we expected. It might be because the RNN model is only concerned with discovering the hidden semantics behind the text and lacks sufficient memory, resulting in keyphrases or words that are too general and do not necessarily refer to the source text. By taking into account more contextual information, our proposed model outperforms not only the RNN model but also all baselines on the Inspec and SemEval-2010, outperforming the best baselines by more than 22% on average. This result demonstrates the significance of using word embedding as a source text as well as extensive training.

5 Conclusion and Future Works

In this paper, we proposed a Deep Bidirectional Long Short-Term Memory Recurrent Neural Network model with word embedding that can capture hidden context and main topics in input documents using keyphrases. The model makes use of Bi-ability LSTM's to extract long-distance semantic information from the input sequence. Since word embedding representation is critical to the model's success, the proposed architecture successfully copes with variable text lengths and assesses their context without requiring hand-crafted features. Furthermore, this architecture is effective at extracting keyphrases from various types of text and domains. Experiment results on three datasets revealed that the proposed Bi-LSTM model, which accounts for both of these dependencies, has a significant impact on keyphrase extraction performance. In the future, we intend to test additional network architectures, such as the generation of missing keyphrases, as well as to test our algorithm's robustness on larger datasets with an extended version in order to improve its performance.

References

1. Alami Merrouni, Z., Frikh, B., Ouhbi, B.: Automatic keyphrase extraction: a survey and trends. *J. Intell. Inform. Syst.* **54**(2), 391–424 (2019). <https://doi.org/10.1007/s10844-019-00558-9>
2. Liu, Z., Li, P., Zheng, Y., Sun, M.: Clustering to find exemplar terms for keyphrase extraction. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, vol. 1, pp. 257–266. Association for Computational Linguistics (2009)
3. Zha, H.: Generic summarization and keyphrase extraction using mutual reinforcement principle and sentence clustering. In: *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 113–120. ACM (2002)
4. Berend, G.: Opinion expression mining by exploiting keyphrase extraction. In: *Proceedings of the 5th International Joint Conference on Natural Language Processing*. Asian Federation of Natural Language Processing (2011)

5. Dashtipour, K., Gogate, M., Cambria, E., Hussain, A.: A novel context-aware multimodal framework for Persian sentiment analysis. *Neurocomputing* (2021)
6. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist.* **5**, 135–146 (2017)
7. Ferrara, F., Pudota, N., Tasso, C.: A keyphrase-based paper recommender system. In: Italian Research Conference on Digital Libraries, pp. 14–25 (2011)
8. Jiang, X., Hu, Y., Li, H.: A ranking approach to keyphrase extraction. In *Proceedings of the 32nd international ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009*, pp. 756–757. ACM, New York (2009). <https://doi.org/10.1145/1571941.1572113>
9. Hulth, A.: Improved automatic keyword extraction given more linguistic knowledge. In: *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pp. 216–223. Association for Computational Linguistics (2003)
10. Frank, E., Paynter, G.W., Witten, I.H., Gutwin, C., Nevill-Manning, C.G.: Domain-specific keyphrase extraction. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence, IJCAI 1999*, pp. 668–673. Morgan Kaufmann Publishers Inc., San Francisco (1999). <http://dl.acm.org/citation.cfm?id=646307.687591>
11. Merrouni, Z.A., Frikh, B., Ouhbi, B.: HAKE: an unsupervised approach to automatic keyphrase extraction for multiple domains. *Cogn. Comput.* 1–23 (2021). <https://doi.org/10.1007/s12559-021-09979-7>
12. Yang, S., Lu, W., Yang, D., Li, X., Wu, C., Wei, B.: KEYPHRASEDS: automatic generation of survey by exploiting keyphrase information. *Neurocomputing* **224**, 58–70 (2017)
13. Papagiannopoulou, E., Tsoumakas, G.: Local word vectors guiding keyphrase extraction. *Inf Process Manage.* **54**(6), 888–902 (2018)
14. Meng, R., Zhao, S., Han, S., He, D., Brusilovsky, P., Chi, Y.: Deep keyphrase generation. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers*, pp. 582–592. Association for Computational Linguistics (2017)
15. Zhang, Q., Wang, Y., Gong, Y., Huang, X.: Keyphrase extraction using deep recurrent neural networks on Twitter. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing* (2016)
16. Alzaidy, R., Caragea, C., Giles, C.L.: Bi-LSTM-CRF sequence labeling for keyphrase extraction from scholarly documents. In: *The World Wide Web Conference on - WWW 2019*, pp. 2551–2557 (2019). <https://doi.org/10.1145/3308558.3313642>
17. Ye, H., Wang, L.: Semi-supervised learning for neural keyphrase generation. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 4142–4153 (2018). <https://doi.org/10.18653/v1/D18-1447>
18. Willis, A., Davis, G., Ruan, S., Manoharan, L., Landay, J., Brunskill, E.: Key phrase extraction for generating educational question-answer pairs. In: *Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale*, pp. 1–10 (2019)
19. Nikzad-Khasmakhi, N., et al.: Phraseformer: multimodal key-phrase extraction using transformer and graph embedding, pp. 1–15 (2021). <http://arxiv.org/abs/2106.04939>
20. Yang, P., Ge, Y., Yao, Y., Yang, Y.: GCN-based document representation for keyphrase generation enhanced by maximizing mutual information. *Knowl. Based Syst.* **243**, 108488 (2022). <https://doi.org/10.1016/j.knosys.2022.108488>
21. Wang, R., Liu, W., McDonald, C.: Corpus-independent generic keyphrase extraction using word embedding vectors. In: *Software Engineering Research Conference*, vol. 39 (2014)
22. Augenstein, I., Søgaard, A.: Multi-task learning of keyphrase boundary classification. In: *Proceedings of ACL (Volume 2: Short Papers)*, vol. 2, pp. 341–346 (2017)
23. Chen, J., Zhang, X., Wu, Y., Yan, Z., Li, Z.: Keyphrase generation with correlation constraints. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 4057–4066 (2018). <https://doi.org/10.18653/v1/D18-1439>

24. Wang, Y., et al.: Exploiting topic-based adversarial neural network for cross-domain keyphrase extraction. In: 2018 IEEE International Conference on Data Mining (ICDM), vol. 2018, pp. 597–606 (2018). <https://doi.org/10.1109/ICDM.2018.00075>
25. Basaldella, M., Antolli, E., Serra, G., Tasso, C.: Bidirectional LSTM recurrent neural network for keyphrase extraction. In: Serra, G., Tasso, C. (eds.) IRCDL 2018. CCIS, vol. 806, pp. 180–187. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73165-0_18
26. Cho, K., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint [arXiv:1406.1078](https://arxiv.org/abs/1406.1078) (2014)
27. Santosh, T.Y.S.S., Sanyal, D.K., Bhowmick, P.K., Das, P.P.: DAKE: document-level attention for keyphrase extraction. In: Jose, J.M., et al. (eds.) ECIR 2020. LNCS, vol. 12036, pp. 392–401. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-45442-5_49
28. Wu, H., Ma, B., Liu, W., Chen, T., Nie, D.: Fast and constrained absent keyphrase generation by prompt-based learning. Proc. AAAI Conf. Artif. Intell. **36**(10), 11495–11503 (2022). <https://doi.org/10.1609/aaai.v36i10.21402>
29. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**, 2493–2537 (2011)
30. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
31. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
32. Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to forget: continual prediction with LSTM. Neural Comput. **12**(10), 2451–2471 (1999)
33. Hulth, A.: Improved automatic keyword extraction given more linguistic knowledge. In: Proceedings of Conference on Empirical Methods in Natural Language Processing (2003)
34. Bouguin, A., Boudin, F., Daille, B.: Topicrank: graph-based topic ranking for keyphrase extraction. In: Proceedings of International Joint Conference on Natural Language Processing (2013)
35. Meng, R., Zhao, S., Han, S., He, D., Brusilovsky, P., Chi, Y.: Deep keyphrase generation. In: Proceedings of the ACL, pp. 582–592 (2017)
36. Mihalcea, R., Tarau, P.: TextRank: bringing order into text. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp. 404–411 (2004)
37. Witten, I.H., Paynter, G.W., Frank, E., Gutwin, C., NevillManning, C.G.: KEA: Practical automatic keyphrase extraction. In: Proceedings of the Fourth ACM Conference on Digital Libraries, pp. 254–255. ACM (1999)
38. Kim, S.N., Medelyan, O., Kan, M.-Y., Baldwin, T.: Semeval-2010 task 5 : Automatic keyphrase extraction from scientific articles. In: Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval@ACL 2010, 15–16 July 2010, pp. 21–26. Uppsala University, Uppsala, Sweden (2010)
39. Christopher, D.M., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The stanford CoreNLP natural language processing toolkit In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55–60 (2014)

Machine Learning and Applications



Feedforward Neural Network in Cancer Treatment Response Prediction

Hanan Ahmed¹(✉), Howida A. Shedeed¹, Safwat Hamad¹, and Ashraf S. Hussein²

¹ Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
hanan.ahmed20100@cis.asu.edu.eg, hanan.ahmed20100@gmail.com

² King Salman International University, El Tur, Egypt

Abstract. Recently, research efforts are exerted on cancer treatment prediction based on the biomarkers related to the tumor. In order to save time and human effort, computational power is used to analyze huge sequences such as human DNA (DeoxyriboNucleic Acid). The use of machine learning and deep learning algorithms becomes a must. Using machine learning (ML) and deep learning (DL) algorithms to predict cancer treatment or drug response is considered a recent approach. There is no specific approach that proves its efficiency against all other approaches, but the indicators show the performance of deep learning-based approaches overcomes others. In this paper, nine different feedforward network architectures are introduced to predict drug responses. The proposed architectures are different in the number of layers and the number of nodes in each layer. Principal Component Analysis (PCA) is used as a dimension reduction method with different degrees of reduction to find the best degree of reduction. The proposed architectures are used with 19702 genes as input to predict the response to 265 different anti-cancer drugs. The proposed Feedforward architectures achieve evolution accuracy over other feedforward model architectures. The enhancement is between 45% and 52% reduction in the Mean Squared Error (MSE).

Keywords: Artificial intelligence · Artificial neural networks · Biomedical · Feedforward neural networks · Personalized medicine · Drug response prediction

1 Introduction

In 2018, there were 18.1 million new cancer cases and 9.5 million cancer deaths, which is considered the second cause of death that year [1]. The expectation says that by 2040, the number of cancer cases will be 29.5 million, and the number of cancer deaths will be 16.4 million [2]. Therefore, immense efforts are exerted to find an effective cancer treatment.

Cancer is abnormal cell growth. There are more than 100 types of cancer; most of them are called by their organs or tissues. Most of the cancer types form some solid tumors, but others do not as leukemia. With the massive variation in cancer characteristics, scientists found that abnormal cell growth is related to gene mutations. The same tumor may contain thousands of mutations; some are drivers that cause cancer, and others are passengers that result from cancer development. With the immense development

in genetics research, they find that some drugs give good results for patients and others do not due to the difference in their genetic profiles and the genetic mutation related to their cases [3, 4]. That provides evidence that the drug response prediction process should be based on a genetic profile with its mutation profile too.

Studying the genetic profiles of each patient and deciding the best treatment is a very time-consuming process and requires a massive human effort. According to [5], machine learning (ML) algorithms were used since 2010 to predict the probability of developing, redeveloping, and survivability of cancer. Since 2017, both ML and deep learning (DL) algorithms have been used for predicting cancer treatment responses.

Different techniques are used for predicting the treatment response. Deep Neural networks (DNN) are used in [6–8]. In [7], a comparison is held among DNN, Random Forest, and Elastic Net models using different datasets. The DNN model outperforms other models on different datasets. In [8], a DL model is developed (consDeepSignaling). It is constrained with 45 pathways and uses gene expression and copy number variation for genes as features. The Convolutional Neural Network (CNN) models are used in drug response prediction, such as in [9, 10]. Cancer Drug Response profile scan (CDRscan) is introduced in [9]. CDRscan consists of five different CNN-based models. A twin Convolutional Neural Network (tCNN) is proposed in [10]. In [11], DNN is introduced for drug response prediction which uses gene expression as input. DNN consists of three phases: pre-training feature selection, using the support vector regression based on recursive feature elimination (RFE), continuous drug response prediction, with a multiple layer DNN, and sensitivity discretization, with the interquartile range (IQR).

In this paper, we propose different feedforward architecture designs for prediction models. The prediction is based on both the gene expression profile and the mutation profile. The used datasets are Cancer Genome Atlas (TCGA), Cancer Cell Line Encyclopedia (CCLE), and Genomics of Drug Sensitivity in Cancer (GDSC). The rest of the paper is organized as follows: Sect. 2 introduces the proposed methodology and its results in Sect. 3. Finally, the comparative study is introduced in Sect. 4, and the conclusion is in Sect. 5.

2 Proposed Methodology

The proposed models are implemented to predict the half-maximal inhibitory concentration (IC_{50}) values. IC_{50} is a quantitative measure that indicates how much of a particular inhibitory substance (e.g. drug) is needed to inhibit, in vitro, a given biological process or biological component by 50%. The drug response data in GDSC is given in form of IC_{50} , so the predicted results will be comparable to the real one. The proposed models predict IC_{50} based on the genomic profiles of a cell or a tumor. The methodology is divided into the following steps:

1. The first part is the dimension reduction to the input data that includes two different Principal Component Analysis (PCAs) systems: one for mutations profile and the other for gene expression profile. Different numbers of PCAs architectures are used to find the best reduction degree.
2. The second part is the prediction network which predicts the IC_{50} values of 265 drugs and uses the output of the two PCAs as an input.

2.1 Datasets

In the proposed work, the used datasets are Cancer Genome Atlas (TCGA), Cancer Cell Line Encyclopedia (CCLE), and Genomics of Drug Sensitivity in Cancer (GDSC). CCLE project aims to accurately characterize the genetic characteristics of cancer cell lines. CCLE includes mutation status for 25 oncogenes across 486 cancer cell lines, DNA copy number variations for 23,316 genes across 1,043 cancer cell lines, and mRNA expressions for 54,675 mRNAs across 127 cancer cell lines [12]. In 2019, the CCLE database received a major update, including newly released DNA methylation data, whole genome sequencing data, and RNA-seq data [13]. The TCGA dataset is comprised of 2.5 petabytes of data with the genomic profiles of tumors and matched normal tissues from more than 11,000 patients representing 33 types of human cancers [14]. The GDSC [15] project provides drug response gene expression from 1,001 cancer-cell lines and drug response data in the form of IC_{50} values for cancer genomic data for 265 drugs. Those drugs are the anti-cancer drugs the GDSC mentions their response to different CCLE cell lines. The numbers of transcripts per million (tpm) for genes are used from CTD² Portal [16]. For TCGA, the Pan-Cancer dataset was used, and tpm values were downloaded from UCSC Xena [17, 18] for 10536 TCGA samples. The data preparation, integration, and federation are used as in [6] to prepare, integrate and improve the data quality. The used samples from both CCLE and GDSC are 619 samples while TCGA is 9158 samples. The number of genes for both CCLE and TCGA is 19702 genes.

Data Federation. Data federation is implemented as introduced in [6]. The datasets that will be federated are CCLE gene expression files from CTD2 Portal, MAF files from CCLE, and CCLE drug response from GDSC. Hence, after the federation, each cell line will contain gene expression data, its own mutation profile, and its drug response data. The input to the data federation are:

1. CCLE gene expression TPM.
2. The mutation file for CCLE.
3. The drug response data from GDSC.

The key to match those inputs are:

1. CCLE gene expression TPM file name (*analysis id*).
2. *DepMap ID* from mutation file.
3. *COSMIC_ID* for drug response data in GDSC

The keys for mapping are:

1. “CCLE_id_mapping.txt” from CTD². It maps analysis id to ccle_name
2. “sample_info.csv” from GDSC. It maps ccle_name to DepMap ID and COSMIC_IDs to their corresponding DepMap IDs.

Not all DepMap IDs have COSMIC_ID so the DepMap IDs that do not have COSMIC_ID are neglected.

2.2 Dimension Reduction

Dimension reduction is the transformation of data from a high-dimensional space into a lower-dimensional space so that the lower representation retains some meaningful properties of the original data. It is common in fields that deal with large numbers of observations and/or large numbers of variables, such as signal processing, speech recognition, neuroinformatics, and bioinformatics. Dimensionality reduction can be used for noise reduction, data visualization, cluster analysis, or as an intermediate step to facilitate other analyses. The most commonly used methods for dimension reduction are principal component analysis (PCA) and auto-encoder.

Auto-encoder is an unsupervised DL (feedforward) that includes symmetric pair of encoder-decoder. It is a recent dimensional reduction method depending on minimizing the loss between the input and reconstructed data (decoded). Principal Component Analysis is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new transformed features are called the Principal Components. PCA is faster and computationally cheaper than auto-encoders. The Auto-encoder is prone to overfitting due to the high number of parameters. As both methods proved their efficiency, PCA is selected due to its run-time and to avoid the probability of the overfitting.

Principal Component Analysis (PCA). Principal Component Analysis (PCA) is a dimension reduction method that is commonly used for the dimension reduction of large datasets. PCA depends on constructing the covariance matrix among all possible pairs of variables, as in (1) and (2), and then Eigenvectors and eigenvalues will be computed. The Eigenvectors represent the principal vectors, and the eigenvalues are the amount of variance. Principal components represent the directions of the data that explain a maximal amount of variance. The maximal amount of variance means the more information it has.

$$x_i^* = (x_i - \mu_i) / \sigma_i \quad (1)$$

$$C = \text{Cov}(X) \quad (2)$$

where x is a data item in n space $x = (x_1, x_2, \dots, x_n)$, $i \in [1, n]$, μ_i is the mean of dimension i , σ_i is the standard deviation for the same dimension and $\text{Cov}(X)$ is $(n \times n)$ covariance matrix.

In the proposed work, two PCA systems are used; one trained on gene expression data and the other one on mutation. The two systems are trained on TCGA data after performing data integration between gene expression data and mutation data, as in [6], (Refer to Fig. (1)).

Various numbers of PCAs are used 50, 64, and 150. As both [20] and [6] uses 64 PCA and its needed to prove if this the best reduction or not 50 is selected to be the less value and to grantee not to lose a lot of information. To study the effect of increasing PCAs, 150 is selected to be more than the double of the reference value 64.

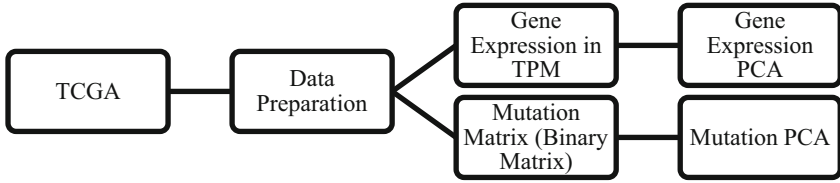


Fig. 1. The Training Sequence of PCA on TCGA dataset

2.3 Feedforward Network-Based Model

The feedforward neural network is the first and the simplest type of artificial neural network. In this network, the information moves in only one direction forward from the input nodes, through the hidden nodes, and to the output nodes, see (Fig. 2). There are no cycles or loops in the network. Feedforward neural networks are primarily used for supervised learning in cases where the learned data is neither sequential nor time-dependent. The feedforward neural networks compute a function f on fixed size input x such that $f(x) \approx y$ for training pairs (x, y) .

Feedforward networks use raw data as input. It can deal with data as it is, but due to the high dimensionality of data, data reduction techniques are used. Both auto-encoders, as in [19], and Principal Component Analysis (PCA) are used. Feedforward is used for the prediction of drug response as in [6, 19].

The Prediction Network. Feedforward is a general-use model which has no feedback connections. The deeper the network is, the better learning and accuracy are, so different architectures are used with different numbers of nodes in each layer. Nine different architectures are introduced; three of them are with 5 layers, and each of those three has a different number of nodes in each layer (100, 128, and 300). The next three are with 10 layers, and the numbers of nodes in each layer for those architectures are 100, 128, and 300. The last 3 are 16 layers and the number of nodes for each architecture is the same. Table 1 shows the summary of the network's structure. The used activation function is the ReLU, except for the output which is the linear activation function. The Adam optimizer is utilized with Mean Square Error (MSE) as a loss function. He's uniform distribution [19] is used to initialize weights. The Feedforward based models use PCA with different degrees of reduction: 50, 64, and 150 principal components. The PCA is trained with TCGA datasets after integration between gene expression data and mutation data, as in [6]. Two PCA systems were utilized: one for the gene expression profiles and the other for mutation profiles. The output of the two PCAs are the input for the prediction network, see Fig. 3.

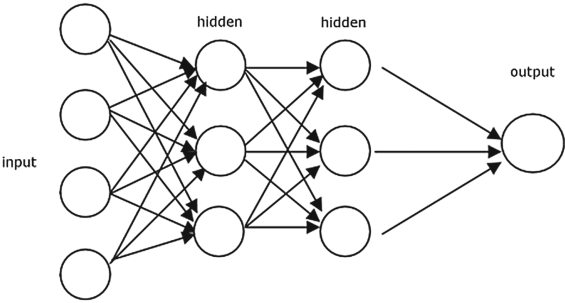


Fig. 2. Feedforward network architecture

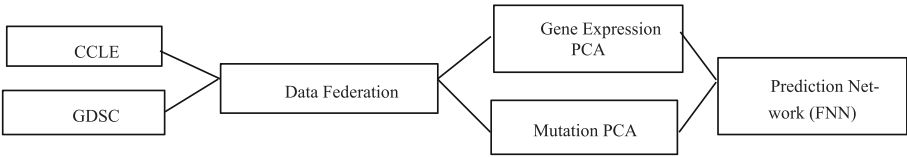


Fig. 3. The sequence diagram for the proposed Model

Table 1. Feedforward models architectures summary

Model name	Number of layers	Number of nodes in each layer
FFN1	5	128
FFN2	5	100
FFN3	5	300
FFN4	10	128
FFN5	10	100
FFN6	10	300
FFN7	16	128
FFN8	16	100
FFN9	16	300

3 Experimental Results

The proposed approach aims to predict the drug response by predicting the IC_{50} . IC_{50} is a half-maximal inhibitory concentration for a drug. IC_{50} is a quantitative measure that indicates how much of a particular inhibitory substance (e.g. drug) is needed to inhibit, in vitro, a given biological process or biological component by 50%. IC_{50} prediction is performed using the genetic and mutation profiles for CCLF and GDSC datasets. The GDSC for CCLF cell lines provides authentic values. CCLF has 619 cell lines for

25 tissues with 19702 genes for each gene expression profile and mutation profile after federation [6].

In Table 2, Feedforward networks with PCA as dimension reduction method results are shown. PCA reduces the input to 64 dimensions only. Two PCAs are used: one for gene expressions and one for mutations. The feedforward Network input is 128 Architecture; FFN7 achieves the best accuracy with MSE less than the MSE of FFN1 and the MSE of FNN4, but with the worst convergence rate, FNN1, FNN4, and FNN7 have 128 nodes in each layer, while FNN7 is the deepest design (16 layers).

Table 3 explains the results for the Feedforward networks using PCA as a dimension reduction method. PCA reduces the input to 50 dimensions only. So, the feedforward Network input is 100. The architecture FNN8 shows the best accuracy with MSE less than FFN2 and FNN5, also its worst convergence rate. FNN2, FNN5, and FNN8 have 100 nodes in each layer, but FNN8 is the deepest design (16 layers).

Table 4 shows the results of feedforward networks with PCA as a dimension reduction method. PCA reduces the input to 150 dimensions only, so the feedforward input is 300. The architecture FNN9 shows the best accuracy with MSE equals 1.7749, and also its best convergence rate. The FNN9 achieves better performance than FNN3 and FNN6. FNN3, FNN6, and FNN8 have 300 nodes in each layer whilst FNN9 is the deepest design (16 layers).

From Tables 2, 3, and 4, the FFN8 has the best performance which indicates that the PCA with 50 dimensions reduction and eliminates the noisy features. Also, FFN7, FFN8, and FNN9 achieve the best performance which indicates that the deepest the network is, the better performance is, as all of them with 16 layers depth.

Table 2. Performance summary for feedforward networks with PCA dimension reduction 64

Model	Best -MSE				Average - MSE			
	Training	Validation	Test	# Epoch	Training	Validation	Test	# Epoch
FFN1	1.8097	1.8030	1.5268	10	1.7767	1.8044	1.8981	8.49
FFN4	1.8059	1.81005	1.5011	6	1.7772	1.7853	1.7918	8.2
FFN7	0.69468	0.81596	0.73056	49	0.82267	0.98803	0.9721	23.88

Table 3. Performance summary for feedforward networks with PCA dimension reduction 50

Model	Best -MSE				Average - MSE			
	Training	Validation	Test	# Epoch	Training	Validation	Test	# Epoch
FFN2	1.8159	1.7007	1.5194	10	1.7775	1.7937	1.7846	7.44
FFN5	1.8190	1.7964	1.4452	7	1.7813	1.7673	1.7823	8.09
FFN8	0.7597	0.8287	0.7544	6	0.8184	0.9792	0.9582	21.3

Table 4. Performance summary for feedforward networks with PCA dimension reduction 150

Model	Best -MSE				Average - MSE			
	training	Validation	Test	# Epoch	Training	Validation	Test	# Epoch
FFN3	1.8181	1.7736	1.4834	8	1.7787	1.8606	1.840	21.02
FFN6	1.8222	1.6863	1.54816	7	1.7807	1.76396	1.8891	8.05
FFN9	1.5415	1.6984	1.5399	4	1.7725	1.7789	1.7749	7.58

4 Comparative Study

From the experimental results, the FFN8 architecture achieves the best performance in terms of accuracy. In this section, a comparative study is performed to compare FFN8 with some recent models to prove its efficiency in terms of accuracy, and its advantage due to its simplicity. Some of the recent models based on DNN are selected, such as Enhanced Deep-DR [6], Deep-DR [20], and CDRscan [9].

Each model's datasets, data features, number of drugs, and even prediction methods are summarized in Table 5. All models work on GDSC as a reference for drug response data. CDRscan uses both gene mutation and SMILES as features. The SMILES is the simplified molecular input line entry specification for the drug that is a feature related to the chemical compounds of the drugs. The maximum number of drugs in consideration is 265 drugs used by Enhanced Deep-DR, Deep-DR, and FFN8. The maximum number of genes in consideration is 28,328 genes used by CDRscan. All models are feedforward models except CDRscan which is the CNN model. All models use Gene expression as a feature and GDSC as the dataset for drug response.

Table 6 shows the comparison among Enhanced Deep-DR, Deep-DR, and FFN8 performance in terms of MSE. FNN8 shows the lowest MSE which indicates higher accuracy and better prediction capability. Table 7 compares in terms of Root Mean Square Error (RMSE) which is the square root of MSE. In Table 7, a comparison between CDRscan and FFN8 indicates that the performance of FFN8 is better than CDRscan,

Table 5. Compares the setting of different algorithms

Algorithm	Datasets	Features	Number of genes	Number of drugs	Model type
Enhanced Deep-DR [6]	CCLE, GDSC, TCGA	Gene expression and mutation	19702	265	FFN
Deep-DR [20]	CCLE, GDSC, TCGA	Gene expression and mutation	15363	265	FFN
CDRscan [9]	CCLP, GDSC	SMILES, gene mutation	28,328	244	CNN
FFN8	CCLE, GDSC, TCGA	Gene expression and mutation	19702	265	FFN

due to the RMSE value. The two tables show that the FFN8 is the best performance in means of accuracy with minimum error.

Table 6. Compares among different models in terms of MSE

Algorithm	MSE
Enhanced Deep-DR [6]	1.78
Deep-DR [20]	1.98
FFN8	0.9582

Table 7. Compares among different models in terms of RMSE

Algorithm	RMSE
FFN8	0.9789
CDRscan [9]	1.069

5 Conclusion

In this paper, nine different feedforward network architectures are introduced to predict drug responses. PCA algorithm is used as a dimension reduction method. Feedforward architectures, FNN7 and FNN8 achieve evolution accuracy over the other FNN architectures. The MSE values are reduced with percentage 45% and 46% respectively. Some of the recent models based on DNN are selected to compare with, such as Enhanced Deep-DR, Deep-DR, and CDRscan. A comparison between CDRscan and FFN8 indicates the performance of FFN8 is better than CDRscan, due to the RMSE value. The maximum number of drugs in consideration is 265 drugs used by Enhanced Deep-DR, Deep-DR, and FFN8. The maximum number of genes in consideration is 28,328 genes used by CDRscan.

The FNN8 achieves higher accuracy in predicting the drug response than Enhanced Deep-DR and Deep-DR. The enhancement over Deep-DR and Enhanced Deep-DR are 45% and 52% reduction in the MSE, respectively.

References

1. World Health Organization: 12 September 2018. <https://www.who.int/news-room/fact-sheets/detail/cancer>
2. Cancer Statistics: NCI. <https://www.cancer.gov/about-cancer/understanding/statistics>. Accessed May 2022
3. Liu, J., Wu, Y., Ong, I., Page, D., Peissig, P., et al: Leveraging Interaction between Genetic Variants and Mammographic Findings for Personalized Breast Cancer Diagnosis. AMIA American Medical Informatics Association (2015)

4. Verma, M.: Personalized medicine and cancer. *J. Personal. Med.* **2**(1), 1–14 (2012)
5. Ahmed, H., Hamad, S., Shedeed, H.A., Saad, A.: Review of personalized cancer treatment with machine learning. In: ICCI (2022)
6. Ahmed, H., Hamad, S., Shedeed, H.A., Saad, A.: Enhanced deep learning model for personalized cancer treatment. *IEEE Access* **20**, 106050–106058 (2022)
7. Sakellariopoulos, T., Vougas, K., Narang, S., Petty, R., Tsirigos, A., Gorgoulis, V.G.: A Deep Learning Framework for Predicting Response to Therapy in Cancer. *Cell Reports* (2019)
8. Zhang, H., Chen, Y., Li, F.: Predicting anticancer drug response with deep learning constrained by signaling pathways. *Front. Bioinform.* **1** (2021)
9. Chang, Y., et al.: Cancer Drug Response Profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Scientific Reports*, pp. 1–11 (2018)
10. Liu, P., Li, H., Li, S., Leung, K.: Improving prediction of phenotypic drug response on cancer cell lines using deep convolutional network. *BMC Bioinform.* **20** (2019)
11. Zhao, Z., Li, K., Toumazou, C., Kalofonou, M.: A computational model for anti-cancer drug sensitivity prediction. In: *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1–4 (2019)
12. Barretina, J., et al.: The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012)
13. Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R.: Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019)
14. TCGA Research Network: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>
15. Yang, W., Soares, J., Greninger, P., Edelman, E.J., et al.: Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41** (2013)
16. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C.: Accurate, fast, and model-aware transcript expression quantification with Salmon. *bioRxiv* (2016)
17. UCSC Xena datasets: University of California, Santa Cruz. <https://xenabrowser.net/datapages/>
18. Goldman, M.J., Craft, B., Zhu, J., Haussler, D.: Abstract 250: UCSC Xena for the visualization and analysis of cancer genomics data. In: *The American Association for Cancer Research Annual Meeting*. Philadelphia (2021)
19. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human level performance on imagenet classification. In: *Proceedings of the IEEE International Conference on Computer Vision* (2015)
20. Chiu, Y., et al.: Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med. Genomics* **12**, 143–189 (2019)



A Genetic Algorithm Approach Applied to the Cover Set Scheduling Problem for Maximizing Wireless Sensor Networks Lifetime

Ibtissam Larhlimi^(✉), Maryem Lachgar, Hicham Ouchitachen, Anouar Darif, and Hicham Mouncif

Department of Mathematics and Informatics, Polydisciplinary Faculty,
University Sultan Moulay Slimane, Beni-Mellal, Morocco
{ibtissam.larhlimifpb,maryem.lachgarfpb}@usms.ac.ma,
h.ouchitachen@gmail.com, anouar.darif@gmail.com, h.mouncif@usms.ma

Abstract. Wireless sensor networks (WSN) are an evolving research area with many applications. In most utility applications, sensor nodes have limited power resources. Therefore, it is crucial to control energy consumption to extend network lifetime effectively. The challenge is to program sensor availability and activity for maximizing network coverage lifetime. This problem is classified as an NP-hard and is called the Maximum Coverage Set Scheduling Problem (MCSS). In this paper, we propose a genetic algorithm for extending WSN lifetime. The proposed algorithm was compared with Greedy-MCSS and MCSSA algorithms. The simulation results show the gain and the good impact by using the genetic algorithm in our approach.

Keywords: Wireless sensor network · Lifetime · Coverage · Genetic algorithm · Cover set scheduling

1 Introduction

Today, innovation is developing quickly. Sensors were initially only employed to measure and keep track of a level's physical value. WSN were created as a result of the development of the internet and research into wireless technologies, which enabled these sensors to be connected [1]. In this field of network, the major challenge is the energy consumption optimization [2]. This constraint depends directly on the lifetime maximization which is the subject of this work.

Wireless sensor networks (WSNs) support the Internet of Things (IoT) infrastructure. The IoT can collect, analyze and deploy a considerable amount of data that will be converted into meaningful information and knowledge. It can connect and communicate via the internet with almost any heterogeneous object in the real world.

The main objective of the proposed research is to maximize the network lifetime when the sensors are randomly distributed over the coverage area of a set

of targets. When there are many sensors, their coverage areas overlap, so not all sensors must detect all targets simultaneously. Many sensors offer a temporary disable the option to save battery power and extend their coverage time. This paper proposes a method to form subsets of sensor nodes that are successively activated for defined durations while the other nodes are in sleep mode, maximizing the duration. This problem, also known as the Maximum Lifetime Coverage Problem (MLCP), is NP-hard and has been extensively studied without considering the energy consumption of the sensors during active operation [3, 4].

This paper focuses on scheduling coverage problems, which are vital in all coverage problems. There are many coverage sets, each consisting of sensors that can completely cover the coverage area. The maximum coverage set scheduling (MCSS) problem aims at finding the optimal scheduling strategy for the coverage set in order to maximize the network lifetime.

The rest of this article is organized as follows. Section 2 surveys the related work. Section 3 presents the definition and formulation of the MCSS problem. Section 4 introduces the proposed genetic algorithm approach. The simulation results are illustrated in Sect. 5. Finally, Sect. 6 concludes this article.

2 Related Works

The coverage in the WSN is often considered a critical performance and quality of service measure [5]. According to [6], it is, therefore, necessary to ensure that an event occurring at any point in the area monitored by the sensors is detected by at least one particular sensor.

There are three typical types of coverage in Fig. 1: a) Target(point) coverage, b) Barrier coverage and c) Area coverage [7].

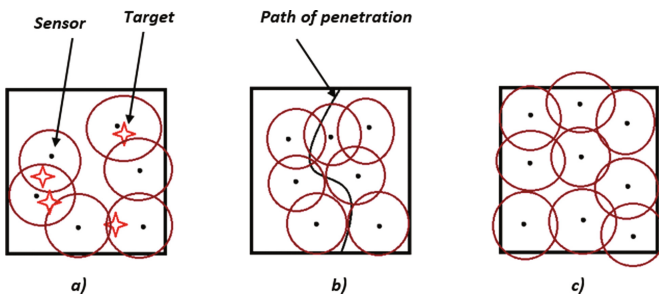


Fig. 1. Types of coverage.

2.1 Target Coverage

Applications in the WSN category known as “target coverage” need the sensor nodes to continuously monitor several distinct points of interest, or “targets” [3].

Due to their limited energy capacity, sensor nodes must be managed to ensure the network has the maximum lifetime. The maximum lifetime coverage problem (MLCP) involves planning the ideal subsets of active or dormant sensor nodes to maximize the network lifetime.

In [8], Lu and al. investigated the area coverage planning problem (MLCS) on a Euclidean plane considering target coverage and data acquisition and proposed a constant factor polynomial-time approximation algorithm for this problem.

In [9], the authors studied the MLCP problem considering the energy consumption of the sensors in standby mode. To solve the MLCP problem, they proposed a new iterative method based on the sensor level and the blacklist concept, which is a binary version of the relationship between a sensor and critical targets.

2.2 Area Coverage

Area coverage is the type of coverage that WSNs employ the most. The latter is always addressed for contiguous space. The coverage can be complete or partial, depending on the application's requirements. Therefore, transforming the area coverage problem into a discrete space problem is usually difficult, which complicates the solution to the area coverage problem.

Tezcan and Wang proposed an algorithm for determining the coverage of a sensor and finding the sensor orientation to maximize the free coverage area for multimedia networks with directional sensors [10].

Consider the costs of heterogeneous sensor deployments and the total budget for a monitored region. Wu and al. developed a new "full-view coverage" model that requires selecting the minimum number of sensors to provide coverage for a given region of interest and proposed a solution for each of two different deployment strategies [11].

In [12], they are interested in solving this problem by using a dynamic planning algorithm to increase the coverage area and cover the hole positions; the dynamic planning algorithm is based on tracking the route to find the optimal path with the most negligible cost.

2.3 Barrier Cover

By forming a conceptual barrier with sensors, barrier coverage is intended to prevent undetected intrusions. The barrier coverage can be categorized as weak or solid [13].

Kumar and al. first introduced the concept of barrier coverage. They then defined the k-barrier coverage notation and proposed algorithms to determine whether k-barrier coverage covers or not a belt area in sensor networks [14].

In mobile sensors, li and Shen studied the problem of barrier coverage with these mobile sensors in a two-dimensional environment to form a specified line segment barrier while minimizing the maximum sensor movement [15].

Wang and al. have considered the barrier coverage problem for sensor position errors and presented the problem of formulating a barrier coverage with fault tolerance to establish a barrier with a guarantee [16].

3 Modelling and Problem Formulation

According to [17], the definition of WSN lifetime is $Lifetime = t_f - t_s$, where t_s is the period when the WSN initiates to operate, and t_f is the period when the network is no longer covered.

The coverage issue can be separated into two critical phases for easier comprehension: Find as many coverage sets from the network sensors as possible first. Second, arrange the obtained coverage sets to extend the network's lifetime.

Consequently, in all coverage problems, the maximum coverage set scheduling problem (MCSS) is crucial. Each coverage set $C_j \in C$ consists of several sensors with an initial monitoring time and can completely cover the coverage area for a given coverage set $C = (C_1, \dots, C_m)$. The MCSS problem is for identifying the optimal scheduling strategy for the coverage set C to maximize the network lifetime [18].

3.1 Mathematical Model

The MCSS can be formulated as an integer linear programming (ILP) problem [18], where $S = \{s_1, s_2, \dots, s_n\}$ is the set of n sensors, for each sensor $s_i \in S$ has b_i active time slots, T_j is defined as the active time of the cover set C_j , and the binary variable $a_{i,j}$ is defined as follows:

$$a_{i,j} = \begin{cases} 1, & \text{if } s_i \in C_j \\ 0, & \text{otherwise} \end{cases}$$

where i : index of sensors, where $1 \leq i \leq n$ and j : index of the cover sets, where $1 \leq j \leq m$.

The MCSS problem is formulated as follows :

$$\max \sum_{j=1}^m T_j \tag{1}$$

Subject to:

$$\sum_{j=1}^m (a_{i,j} T_j) \leq b_i, \forall s_i \in S \tag{2}$$

The overall active time for the C collection of cover sets is to be increased by objective function 1. The total active time intervals in the scheduling strategy must be fewer than or equal to the initial active time intervals for each sensor $s_i \in S$, according to constraint 2.

4 Proposed Approach: Genetic Algorithm

This section contains our proposed genetic algorithm for the MCSS problem. The network lifetime will be increased by discovering a scheduling strategy for

the cover sets in C that guarantees just one cover set is active during each time slot and maximizes their total active time. We propose a genetic algorithm for MCSS problem as follows:

Coding

The chromosome C can be represented as $C = \{C_1, C_2, \dots, C_m\}$, where m is the number of genes, which is equal to the number of covers $C_j \in C$, and each gene represents the set of covers C_j .

Initialization

The initial population of a limited number of chromosomes is randomly generated. A gene represents the cover set C_j . Therefore, the chromosome represents the scheduling strategy of the collection of cover sets C , the lifetime can be calculated by the sum T_j of the genes in the chromosome.

Fitness

The fitness must ensure that not all sensors in the candidate solution have exceeded the energy constraint. For a sensor $s_i \in S$, the sum of energy consumption across all covers in a schedule must not exceed the initial energy b_i (Sect. 3.1 - Eq. 2).

$$\sum_{j=1}^m (a_{i,j} T_j) \leq b_i, \forall s_i \in S \quad (3)$$

The fitness function will exclude the candidate schedule from the population that does not meet this constraint from future GA processes.

Selection

From the members of the adapted population, select the two best candidate chromosomes with the maximum lifetime as parents ((Sect. 3.1 - Eq. 1)).

$$\max \sum_{j=1}^m T_j \quad (4)$$

Crossover

The crossover operator is applied to produce offspring for the following generation and update the population. Double-point crossovers are used where the points are chosen randomly.

Mutation

The mutation operator is applied to improve the children and avoid the optimal local solution. One or more genes are randomly selected and updated. We consider that enhancement increases the value of the genes and the chromosome.

Fitness of children

The fitness function ((Sect. 3.1 - Eq. 2)) is reapplied for the “new chromosome” children to ensure that no sensor exceeds its initial energy. Parents are updated if the new and improved children have a better gene pool or a better lifespan.

5 Simulation and Results

We simulate a network with N sensors randomly placed in a $1000\text{ m} \times 1000\text{ m}$ area and detecting T targets. We apply the proposed algorithm to these coverage sets to measure the network lifetime. Each test result is an average of 100 runs. The performance of the proposed algorithm was evaluated by comparing it with the Greedy-MCSS and MCSSA presented in [18]. In our simulation, we consider the following parameters (Table 1):

Table 1. Parameters of simulations.

Length of chromosome	The scheduling strategy of the collection of cover sets C
Iteration	100
R (Sensing range of each sensor node)	200
T (Target)	10 to 20
M (Number of coverages sets)	100

Using $T = 10$ targets and $N = 300$ sensors, randomly assigned each sensor's active time slots between 10 and 30, and varied the coverage sets from 10 to 60, as shown in Fig. 2. We can see that the suggested algorithm produces a network lifespan that increases in direct proportion to the number of coverage sets. For example, in a network of 40 cover sets, the proposed approach determines the network lifespan to be 389s. In comparison, the Greedy MCSS and MCSSA algorithms determine the lifetime as 170s and 184s, respectively. The findings demonstrate that the proposed algorithm calculates lifetimes superior to those produced by the Greedy-MCSS and MCSSA algorithms.

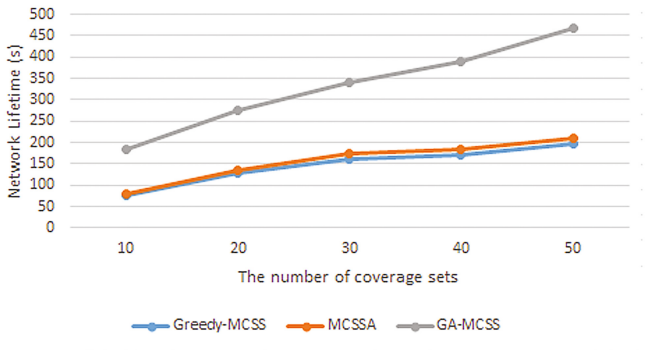


Fig. 2. Varying the number of coverage sets from 10 to 60.

In the second experiment, $M = 100$, $T = 20$ randomly distributed targets, and the number of sensors was varied from 200 to 1000 in increments of 200; we set a time slot of 10 for each sensor. Figure 3 illustrates the approach results as a function of the number of sensors compared to Greedy MCSS and MCSSA. A network lifetime increases in proportion to the number of sensors. Consequently, the fitness and selection part of the genetic algorithm selects the two best proposed solutions with the maximum lifetime as parents, so the chance of combining the best parents to produce good offspring increases with the number of probes. According to the results, the proposed algorithm calculates lifetimes better than Greedy MCSS and MCSSA.

Figure 4 shows the lifetime calculated by the proposed algorithm as a function of the active time slots of the sensors in comparison with the other algorithms. Using $T = 20$, $N = 100$, $M = 100$, change the active time slots from $[0, 10]$ to $[50, 60]$, where $[0, 10]$ means that each sensor's active time slots are randomly assigned from 1 to 10. Based on the results, the proposed approach results in better lifetimes than the two algorithms, Greedy-MCSS and MCSSA, since the proposed algorithm produces offspring with good genomes each time we choose and combine the best parents, resulting in parents who have the maximum lifespan.

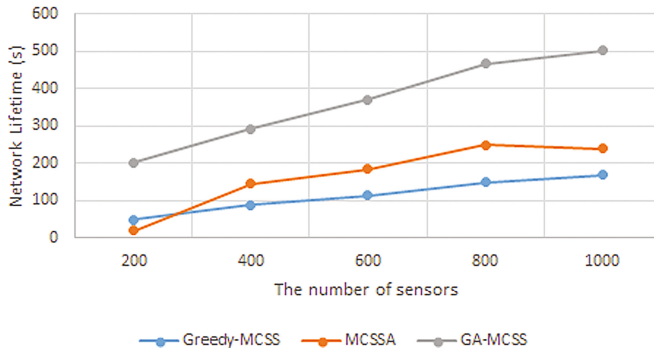


Fig. 3. Varying the number of sensors from 200 to 1000.

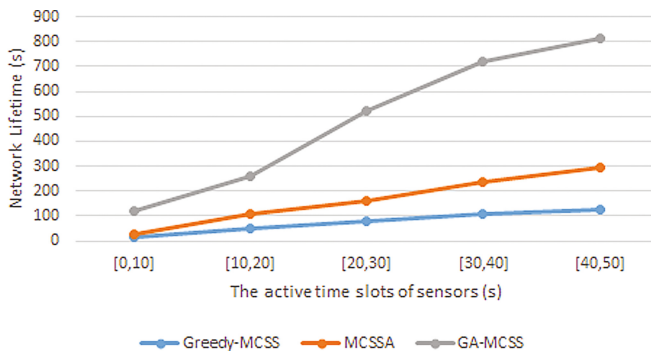


Fig. 4. The active time slots of sensors (s).

6 Conclusion

In this paper, we study the MCSS problem, which involves finding the best optimal coverage and scheduling strategy to maximize the lifetime of WSN. The MCSS problem has been formulated as an integer linear programming problem since it has been proven NP-hard. GA-MCSS is proposed as an algorithm to solve this problem and increase the network lifetime.

The simulation results show that the proposed algorithm provides a more extended network lifetime due to the performances of Genetic algorithm in terms of optimization constrain problems.

In the future, this study can be improved to introduce and analyze more parameters such as crossover, mutation, mobility,...

References

1. Darif, A., Ouchitachen, H.: Performance improvement of a new mac protocol for ultra wide band wireless sensor networks. *J. Theor. Appl. Inf. Technol.* **100**(4), 1015–1026 (2022)
2. Ouchitachen, H., Hair, A., Idrissi, N.: Improved multi-objective weighted clustering algorithm in wireless sensor network. *Egyptian Inform. J.* **18**, 45–54 (2017)
3. Thai, M.T., Wang, F., Du, D.H., Jia, X.: Coverage problems in wireless sensor networks: designs and analysis. *Int. J. Sensor Networks* **3**(3), 191 (2008)
4. Cardei, M.M.T., Thai, Y., Li, W.: Energy-efficient target coverage in wireless sensor networks. In: *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, 1976–1984 (2005)
5. Serper, E.Z., Altın-Kayhan, A.: Coverage and connectivity based lifetime maximization with topology update for WSN in smart grid applications. *Comput. Networks* **109** (2022). <https://doi.org/10.1016/j.comnet.2022.108940>
6. Khoufi, I., Minet, P., Laouiti, A., Mahfoudh, S.: Survey of deployment algorithms in wireless sensor networks: coverage and connectivity issues and challenges. *Int. J. Auton. Adapt. Commun. Syst.* **10**, 314–390 (2017)
7. Amutha, J., Sharma, S., Nagar, J.: WSN strategies based on sensors, deployment, sensing models, coverage and energy efficiency: review, approaches and open issues. *Wirel. Pers. Commun.* **111**(2), 1089–1115 (2020)
8. Lu, Z., Li, W.W., Pan, M.: Maximum lifetime scheduling for target coverage and data collection in wireless sensor networks. *IEEE Trans. Veh. Technol.* **64**(2), 714–727 (2015)
9. Wang, Y., Wu, S., Chen, Z., Gao, X., Chen, G.: Coverage problem with uncertain properties in wireless sensor networks: a survey. *Comput. Networks*, 200–232 (2017). <https://doi.org/10.1016/j.comnet.2017.05.008>
10. Tezcan, N., Wang, W.: Self-orienting wireless multimedia sensor networks for maximizing multimedia coverage. *Comput. Netw.* **52**(13), 2558–2567 (2002)
11. Wu, P.F., Xiao, F., Sha, C., Huang, H.-P., Wang, R.-C., Xiong, N.X.: Node scheduling strategies for achieving full-view area coverage in camera sensor networks. *Sensors* **17**(6) (2017). <https://doi.org/10.3390/s17061303>
12. Alhaddad, Z.A., Manimurugan, S.: Maximum coverage area and energy aware path planner in WSN. In: *Materials Today: Proceedings* (2021). <https://doi.org/10.1016/j.matpr.2020.12.1218>

13. Zhang, X., Fan, H., Lee, V.C.S., Li, M., Zhao, Y., Lin, C.: Minimizing the total cost of barrier coverage in a linear domain. *J. Combin. Optim.* **36**, 434–457 (2018)
14. Kumar, S., Lai, T.H., Arora, A.: Barrier coverage with wireless sensors. In: *Proceedings of the 11th Annual International Conference on Mobile Computing and Networking*, pp. 284–298 (2005)
15. Li, S., Shen, H.: Minimizing the maximum sensor movement for barrier coverage in the plane. In: *IEEE International Conference on Computer Communications (INFOCOM)*, pp. 244–252 (2015)
16. Wang, Z., Chen, H., Cao, Q., Qi, H., Wang, Z., Wang, Q.: Achieving location error tolerant barrier coverage for wireless sensor networks. *Comput. Netw.* **112**, 314–328 (2017)
17. Shi, T., Cheng, S., Cai, Z., Li, J.: Adaptive connected dominating set discovering algorithm in energy-harvest sensor networks. In: *IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1–9 (2016)
18. Chuanwen, L., Yi, H., Deying, L., Yongcai, W., Wenping, C., Qian, H.: Maximizing network lifetime using coverage sets scheduling in wireless sensor networks. *Ad Hoc Networks* (2019)



Application of Machine Learning to Sentiment Analysis

Oumaima Bellar^(✉), Amine Baina, and Mostafa Bellafkih

RAISS Team, STRS Lab, National Institute of Posts and Telecommunications INPT Rabat,
Rabat, Morocco

oumaimabellar@gmail.com, baina@inpt.ma, bellafkih@inpt.ac.ma

Abstract. The Internet is a vast communication tool that has become an indispensable tool for the exchange of information both at personal and professional level; and that allows access to data of all kinds, texts, photos, music, videos, through encryption and universal coding. Due to the rapid increase in the amount of data generated by users on social media platforms such as Twitter, several opportunities and new open doors have been created for organizations that strive to track customer opinions on their products. Sentiment analysis is one of the new challenges appeared in automatic language processing with the advent of social networks. Sentiment analysis is a process that automatically determines whether a user-generated text expresses a positive, negative or neutral opinion about an entity (i.e., product, people, subject, event, etc.). The purpose of this document is to provide detailed steps in the process of analyzing feelings on a data set using machine learning.

Keywords: Sentiment analysis · Machine learning · Opinion mining · Natural language processing (NLP) · Polarity

1 Introduction

Nowadays, social networks have changed the way that people express their opinions and points of view [1]. This occasion is given through textual publications, online discussion spots, product evaluation websites etc. People calculate heavily on this user-generated content. Social networks offer considerable quantum of content generated by the users, it's important content for analysis and offer further services acclimated to the requirements of druggies [8, 9]. In recent times, the maturity of developments in the field of information and opinion exchange have launched the exploration work for the analysis of passions expressed on these social networks, presented in the literature as sentiment analysis [1].

Millions of people are using Twitter and ranked as one of the most visited customers with the normal of 58 million tweets per day [2] However, social networks similar as Facebook, Twitter and Google are decreasingly associated with numerous social marvels similar as importunity, intimidation, and Covid-19. Social networks have attracted the attention of experimenters, which attempts to understand and analyze, among others, the

structure of interconnection and the interaction of users in social networks. Internet users tend to express their opinions and passions and talk about their lives and conditioning of everyday life via Twitter [2]. Thus, it's necessary to automate the system of analyzing feelings to grease the tasks of determining the opinions of the public without having to read numerous tweets manually [5]. This process of analyzing and synthesizing the opinions expressed in the data generated by these enormous users is generally appertained to as sentiment analysis or opinion mining, which is a veritably intriguing and popular area for researchers currently [2].

In 2002, Pang's work on the use of supervised machine learning algorithms to accomplish sentiment classification, such as Naive Bayes, Maximum Entropy, and Vector Machine Support [2], first presented the idea of sentiment analysis and opinion exploration. In the contemporary reality, businesses and associations have consistently learned what the public or customers think about their goods and services [3]. Before deciding on a vote option in a political decision, individual consumers consider their own sentiments as well as that of previous readers of the article they are purchasing and the feeling of others toward political rivals [5]. The opinions voiced may be characterized as either positive or negative, which is helpful to organizations, the government, and individuals in figuring out what the public thinks about a certain issue (product, people, subject, event, etc.) [2].

The objective of the classification of sentiment is to calculate the polarity of the sentences extracted from the text of the review. And let us take the example of the document [2], which shows the effectiveness of the application of machine learning techniques to the problem of classification of feelings to know whether the film is successful or not. This analysis allows the organizations concerned to find people's views on the films and their criticisms, whether positive or negative [5].

The difficult aspect in analyzing sentiments is a word of opinion that is considered as a positive in one situation could also be considered as negative in another situation. Normal word processing considers that a tactile change in two bits of content has no change in meaning or meaning [1]. But in the analysis of sentiments, a small change in two pieces of content has changed in meaning or meaning, let's take the example "film is good" is different from "the film is not good". The system processes it by analyzing one sentence at a time [8]. However, blogs and twitter contain more informal phrases that the user can understand, but the system cannot understand it. In this work, we compute the semantic similarity of a tweet collected from Twitter and training data. We use several machine learning algorithms to build our work. In this context, we focus on COVID-19 pandemic, positive and negative thoughts. We evaluate our method and present the results to predict the semantic orientations of Twitter data.

2 Sentiment Analysis Techniques

Sentiment analysis can be divided into lexicon sentiment analysis technique, machine learning based sentiment analysis, and hybrid methods [9].

2.1 Lexicon Based Approach

It is argued to find the lexicon of opinion to calculate sentiment for a given text. It is a matter of counting the number of positive and negative words in the text [1]. If the text consists of more positive words, the text is assigned a positive rating. If there are more negative words, the text is assigned a negative score.

Otherwise, if the text contains an equal number of positive and negative words, it is assigned a neutral rating [3]. To determine whether a word is positive or negative, an opinion lexicon (positive and negative opinion words) is made [5]. There are several types to compiling and constructing a lexicon of opinion [5] (Table 1):

Table 1. Different types of lexicons based approach.

Type of Lexicon	Description
Sentiment words	Positive or Negative sentiment words have a sentiment score of + 1 or − 1 to indicate the respective polarity
Negation Words	Negation words are the words which reverse the polarity of sentiment
Blind Negation Words	Blind negation words operate at a sentence level and points out the absence or presence of some sense that is not desired in a product feature
Split Words	Split words are the words used for splitting sentences into clauses. The split words list consists of conjunctions and punctuation marks

As each word in the sentence comes into different categories of lexicons, the sentiment can be calculated accordingly. The sentiment score of a sentence constitutes the sum of polarities of each word of the sentence [13].

2.2 Machine Learning Based Approach

Here, two sets of documents serenaded: training-set and test-set. A supervised learning classifier uses the training set to learn and train with respect to the distinctive attributes of the text and the classifier’s performance is tested using a test dataset [5]. Several machine learning algorithms such as Maximum Entropy (ME), Naive Bayes (NB) and Support Vector Machines (SVM) are generally used for text classification. Before performing the analysis of feelings on the data, they must format the data and extract the relevant character is tics of the feelings [2]. Machine learning for sentiment analysis begins with the collection of data sets containing labeled data. These data can be noisy and must therefore be pre-processed using various natural language processing (NLP) techniques [3]. Then the features relevant to the analysis of sentiment were extracted, and the classifier is finally trained and tested on invisible data [5]. These steps are explained in detail in Sect. 4. Machine learning algorithms are further classified in the following categories as explained by [2]. We can use a different type for this approach such as: Supervised Learning, Un-Supervised Learning, Semi-Supervised Learning and Reinforcement Learning.

2.3 Hybrid Approach

In order to improve the performance of the classification of sentiment, few research techniques suggest using a combination of lexical and machine learning techniques. The main advantage of this approach is that we can get the best of both worlds [2].

3 Sentiment Analysis Procedure: Materials and Methods

Prior to conducting the sentiments analysis, the data should be formatted appropriately and the relevant features of the sentiments extracted [18]. To complete this work, follow these steps (Fig. 1):

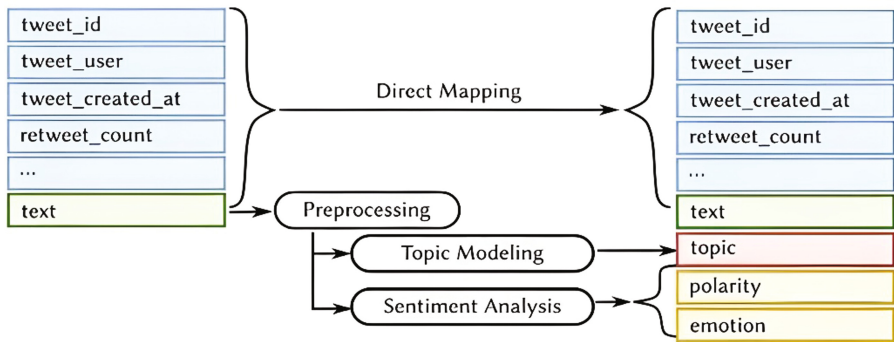


Fig. 1. Detailed natural language processing steps.

3.1 Data Collection

The collection of tweets from Twitter is an essential part of our work. In this sense, Twitter API is used to collect tweets in real time from the Twitter page. In this work, we collect tweets using specific search words related to our theme (Coronavirus - Covid-19). The processing mechanism is used to collect sequential data from Twitter and the extraction is performed several times for more tweets [11]. The tokens and access keys obtained, are necessary for the extraction of tweets in real time from the Twitter page, containing users 'opinions and views about the target entity (product, people, topic, event, etc.) [2].

Opinion Mining or the analysis of sentiment allows us to automatically analyze the textual data and to highlight the different opinions expressed on a specific subject such as a brand, a press article or a product [2]. The proposed approach follows the steps out-lined in Sect. 3 for analyzing feelings on Twitter data. I chose the social network Twitter, which is currently the most popular micro-blogging platform. Online social data contains many real events that have occurred in lifestyle, and the global disease caused by COVID-19 is now spreading around the world. Many people, including the media and government agencies, present the latest news and opinions about the Coronavirus

[2]. Twitter allows researchers to collect tweets by using a Twitter API. One must have a twitter account to obtain twitter credentials (i.e., API key, API secret, Access token and Access token secret) which can be obtained from twitter developer site. Then install a twitter library to connect to the Twitter API. Twitter has developed its own language conventions [3]. Twitter data was collected, through the python programming language with the keywords #COVID-19, #Coronavirus, from August to November 2023. In this project, we have tested several raw data pre-processing techniques from Sect. 3.2, and Naïve Bayes, Logistic Regression and Decision Tree Machine Learning algorithms are used for the analysis of sentiments [12]. This NLP- based model uses the Stanford NLTK with the TextBlob library to create functions that calculate the polarity and subjectivity of tweets. According to the results of our ‘Polarity’ and ‘Subjectivity’ function we can analyze each Tweet in our dataset, and refers to the identification of the sentiment orientation (positive, neutral and negative) and give us a good structuring of the datasets. An excellent style manual for science writers is [7] (Table 2).

Table 2. Number obtained of tweets of each type of sentiment.

Sentiment	No of tweets
Positive	3417
Neutral	2265
Negative	1818

And that’s the percentage of each tweet positive, neutral and negative (Table 3).

Table 3. Percentage obtained of tweets of each type of sentiment.

Sentiment	Positive	Neutral	Negative
Percentage of tweets	0.4556	0.302	0.2424

3.2 Data Pre-processing

In complete pre-processing, data cleaning was performed to improve the learning efficiency of machine learning models. Machine learning models show improved classification accuracy if the data are pre-processed [8]. There are several techniques for data preprocessing, among these techniques are:

Lower Case. The conversion of the data to lowercase letters is one transformation that has been performed to all the data. This is due to the fact that the characters “A” and “a” are represented differently in the machine’s memory and assigned distinct numbers, although we already know that the words “Play” and “play” are interchangeable. The

foundation of all machines learning techniques is numerical computing. As we can see in the next section, changing the data to lower case has both positive and bad implications. For example, “Apple,” which refers to the company, is not the same as “apple” the fruit.

Stop-words. The performance of some issues, such as search engines, can be improved by reducing the number of features in the final vector and eliminating words like connectors that do not have a specific meaning by themselves. But one must keep in mind that leaving out these phrases could result in a different sentence with a different meaning. For example, “not” is considered a stop word but if we delete it in the sentence “I’m not happy” the meaning has changed to the inverse “I’m happy” and make a wrong classification.

Punctuation. Punctuations, such as exclamation (!) or question (?) marks, can be unnecessary and only slightly affect the sentimental meaning of a word. Because a positive or negative sentiment has no bearing on whether something is a question or an appreciation, the dot(.) can also imply that one feeling has ended and a new one is beginning in much the same way.

Stemming. The idea underlying stemming is that a single word has a variety of different morphologies in natural language. Every morphology will appear as a separate word to the computer [7]. Stemming eliminates duplicates of these terms by grouping words together based on a common root. For instance, the verb “argue, argued, argues and arguing” will be reduced to “argue”.

Lemmatization. Strives to use the natural word root or its base form, known as the “lemma,” rather than stemming words. For example, the word “meeting” has the lemma “meet” [12]. This method aims to address potential stemming collisions. The fundamental issue is that there isn’t a dictionary for every term, and informal uses of the words are unknown and can be treated as distinct people. I utilized the Natural Language Toolkit’s parser and dictionary for this purpose [2].

Dictionary of Words. With this method, a dictionary containing all words in the texts is created and then each word in the text is converted to the index of the word inside the dictionary [7]. This method does not break the words order and, also, does not group the words. This transformation can generate a huge dictionary because the machine is case sensitive, for example, and can generate a different index for the same word. Also, is hard to have a dictionary with all words on any language and all of its forms, formal or informal [8].

3.3 Features Extraction Methods

After the pre-processing phase, the data was divided into “training subset” and “testing subset”. It was divided in the ratio of 3:1 for training and testing, respectively. Feature extraction methods were then applied to the training subset [4]. Feature extraction techniques were applied to both training and testing data: on the training data to train the selected models and on the testing data when classification was performed [8]. So, it is necessary to extract the relevant features for sentiment analysis. Some of these features are:

Term Presence and Frequency. Individual words, word n-grams, and their frequency counts make up these features. It either utilizes term frequency weights to represent the relative relevance of aspects [14] or gives terms binary weighting (zero if the word exists, one if it doesn't, etc.).

Opinion Words and Phrases. The content contains words and phrases that express an opinion [4]. These are terms that are frequently used to communicate opinions, such as positive or negative, like or hate. On the other hand, some expressions of opinions can be made without the use of opinion words [6]. As an illustration, “cost me an arm and a leg”.

Negation. Presence of words likes “not”, “nor”, “neither” can reverse the sentiment.

Specific Features. Presence of emoticons in the data, positive or negative hashtags because they add meaning to the sentiment [4].

TF-IDF Method. TF is denoted by term frequency whereas IDF is denoted by inverse document frequency. This method is also dependent on frequency but can be distinguished from the former method, in such a way that it sees for occurrence of a word in the whole corpus [3]. This approach assigns lowest weight to the most repeated words in the corpus and gives preference to words which have occurred very less times in corpus.

The significant phrases used in TF-IDF are:

- **TF** = (the count of the term t appears in a text document)/(Number of terms in the document).
- **IDF** = $\log(D/d)$, where, D is the number of documents and the number of documents a term t has appeared is referred to as d .

$$TF - IDF = TF * IDF \quad (1)$$

3.4 Training and Testing Machine Learning Classifier

After cleaning and selecting the necessary data, a Machine Learning Classifier should be chosen for sentiment analysis since the training set is used to train and educate the classifier and their performance is measured using the test set [5].

4 Machine Learning Algorithms for Sentiment Classification

This section describes the necessary details for the machine learning classifiers used in this study for tweet classification [5]. There are several Machine Learning algorithms for sentiment analysis, among these algorithms we find:

4.1 Logistic Regression Algorithm

Logistic regression is a supervised machine learning technique for classification problems. Supervised machine learning algorithms train on a labeled dataset along with an answer key which it uses to train and evaluate its accuracy. The goal of the model is to learn and approximate a mapping function $f(X_i) = Y$ from input variables $\{x_1, x_2, x_n\}$ to output variable (Y) [7]. It is called supervised because the model predictions are iteratively evaluated and corrected against the output values, until an acceptable performance is achieved [9].

We make use of the Logistic Regression algorithm to build the model [11]. It identifies the probability of occurrence of an event by fitting data to a logit function. The equation used in the algorithm is:

$$\log(p/(1-p)) = \beta_0 + \beta(\text{num}) \quad (2)$$

Here, If the $\log(p/(1-p))$ is greater than zero, then the success ratio is every time appears to be greater than half of the 100% [12].

$$\text{F1-Score} = 2 * (A * B / A + B) \quad (3)$$

This model based on logistic regression that could help us to predict whether the given Tweet is racist or not by using TF-IDF (Term-Frequency-Inverse-Document Frequency) which measures the importance of each word in a document, relative to a collection or corpus. The accuracy of this model comes out to be 66.15% (Table 4).

Table 4. Results obtained using the logistic regression algorithm.

	Sentiment	Positive	Neutral	Negative
Predicted	Percentage of tweets	0.72	0.65	0.57

4.2 Naïve Bayes Classifier (NB)

The Naïve Bayes classifier (multinomial or not) allows probabilistic classifications, which attribute the probability of belonging to a class. The Naive Bayes classifier works well for text classification because it calculates the posterior probability of a class supported word distribution (features) in the document. This algorithm can be used to categorize the sentiments of a text (positive, neutral or negative), or whether an email is spam or not [3]. The basic idea of Naïve Bayes' technique is to find the probabilities of the classes assigned to the texts using the joint probabilities of words and classes [2]. This is a type of generative algorithm, where we model each class and try to determine the probability of an observation belonging to this class [4]:

$$P(\text{label/features}) = P(\text{features label}) * P(\text{label})/P(\text{features}) \quad (4)$$

- P(label): The prior probability of a label or the likelihood that a label is observed.
- P(features | label): The prior probability that feature set is being classified as a label.
- P(features): The prior probability that a given feature set is occurring.

Given the Naive assumption which states that all features are independent of each other, the equation could be rewritten as follows:

$$P(\text{label}|\text{features}) = \frac{P(\text{label}) * P(f_1|\text{label}) * \dots * P(f_n|\text{label})}{P(\text{features})} \quad (5)$$

Naive Bayes classifier work on the principle of probabilities and the Bayes rule given by:

$$p\left(\frac{c}{d}\right) = \frac{p(c)p\left(\frac{c}{d}\right)}{p(d)} \quad (6)$$

where P(c|d) is the probability of a given document (text) belongs to class c, which is the classification part which we are interested in. Below is that the confusion matrix for the Naive Bayes in our project... The classifier has obtained accuracy of 65.51% (Table 5).

Table 5. Results obtained using the naïve bayes classifier.

	Sentiment	Positive	Neutral	Negative
Predicted	Percentage of tweets	0.48	0.37	0.28

4.3 Support Vector Machine (SVM)

The main idea of SVM is to find linear separators or hyper planes in the search space that can best separate the different classes. There may be several hyper planes separating classes, but the one that is chosen is the hyper plane in which the normal distance of the data points is greatest [7], so it represents the maximum separation margin [5]. Text classification is well suited to MVMs due to the sparse nature of the text, in which few features are irrelevant, but they tend to be correlated with each other and generally organized into linearly separable categories [4].

Below is the confusion matrix of the performance of the Support Vector Machine. We can see that this classifier has misclassified a greater number of knowledge points as compared to Naïve Bayes [13]. The accuracy of this model comes out to be 43.63%, which is lower than that for Naïve Bayes (Table 6).

Table 6. Results obtained using the support vector machine algorithm.

	Sentiment	Positive	Neutral	Negative
Predicted	Percentage of tweets	0.48	0.37	0.28

4.4 Decision Trees

The training data space is represented in a hierarchical form in which a condition on the value of the attribute is used to partition the data (the condition of the attribute values is the presence or absence of one or more words) [12]. Data space partitioning is done recursively until leaf nodes contain a minimum number of records used for classification purposes [5]. Decision tree technique is one of the most intuitive and popular in Data Mining, especially as it provides explicit ranking rules and supports heterogeneous, missing data and non-linear effects well. The accuracy of this model comes out to be 60.95% (Table 7).

Table 7. Results obtained using the decision tree algorithm.

	Sentiment	Positive	Neutral	Negative
Predicted	Percentage of tweets	0.69	0.59	0.48

5 Discussion and Future Work

In this study, we describe a number of machine-learning-based techniques for employing Twitter as a preventative in the situation of COVID-19. Our technique also allows for the semantic analysis of Twitter data. We intend to continue to develop and hone our methods in order to increase the precision of our methodology. Following that, well-known algorithms and methods that are typically employed for classifying sentiments are studied. A comparison of precision across several datasets is given, and it might serve as a guide for next studies.

6 Conclusion

Sentiment Analysis can be performed using a lexical approach, a machine learning-based approach or a hybrid approach. The lexicon-based approach faces a drawback that the strength of the sentiment classification depends on the size of the lexicon (dictionary) [13]. As the size of the lexicon increases, this approach becomes more erroneous and time-consuming. This document explains the different steps to perform sentiment analysis on data using Machine Learning algorithms and methods. A Machine Learning

Classifier requires a labeled dataset that is divided into which is divided into training-set and test-set [8]. So, the next step is to perform pre-processing of the data using NLP based techniques and methodologies, followed by a feature extraction method followed by a feature extraction method to extract relevant features [10]. Finally, a model is trained using machine learning classifiers such as Naïve Bayes, Support Vector Machines, Logistic Regression and Decision trees. This model is tested on test data. The performance of the model can be measured in terms of precision, accuracy, recall and F1-score.

References

1. Ahmad, M., Aftab, S., Muhammad, S.S., Ahmad, S.: Machine learning techniques for sentiment analysis: a review. *Int. J. Multi-Sci. Eng.* **8** (2017)
2. Wawre, S.V., Deshmukh, S.N.: Sentiment classification using machine learning techniques. *Int. J. Sci. Res.* (2016)
3. Hasan, A., Moin, S., Karim, A., Shamshirband, S.: Machine Learning- Based Sentiment Analysis for Twitter Account. MDPI (2018)
4. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng. J.* **5**, 1093–1113 (2014)
5. Jain, A.P., Dandannavar, P.: Application of machine learning techniques to sentiment analysis. In: 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) (2016)
6. Llombart, O.R.: Using Machine Learning Techniques for Sentiment Analysis. Final Project on Computer Engineering School of Engineering (EE). Universitat Autònoma de (UAB) (2017)
7. Ferri, A.D.F.: Approaches, tools and applications for sentiment analysis implementation. *Int. J. Comput. Appl.* **125** (2015)
8. Singh, J., Singh, G., Singh, R.: Optimization of Sentiment Analysis Using Machine Learning Classifiers. Department of Computer Science, Guru Nanak Dev University, Amritsar, India Singh et al. *Hum* (2017)
9. Coutinho, D.P., Figueredo, M.A.T.: An Information Theoretic Approach to Text Sentiment Analysis. Instituto Superior de Engenharia de Lisboa (2015)
10. Korani, W., Mouhoub, M.: Sentiment Analysis of Serious Suicide References in Twitter Social Network. Department of Computer Science, University of Regina (2016)
11. Natarajan R., Barua G., Patra M.R.: Distributed Computing and Internet Technology. Lecture Notes in Computer Science (2015)
12. Gonzalez-Marron D., Mejia-Guzman D., Enciso-Gonzalez A.: Exploiting Data of the Twitter Social Network Using Sentiment Analysis. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (2017)
13. Singh, J., Singh, G., Singh, R.: Optimization of Sentiment Analysis Using Machine Learning Classifiers. Department of Computer Science, Guru Nanak Dev University, Amritsar, India (2017)
14. Gonzalez-Marron, D., Mejia-Guzman, D., Enciso-Gonzalez, A.: Exploiting data of the twitter social network using sentiment analysis. In: Sucar, E., Mayora, O., Munoz de Cote, E. (eds.) *Applications for Future Internet*. LNCS, vol. 179. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-49622-1_5



Robust Vehicle Detection by Using Deep Learning Feature and Support Vector Machine

Vinh Dinh Nguyen^(✉), Thanh Hoang Tran, Doan Thai Dang, and Narayan C. Debnath

School of Computing and Information Technology, Eastern International University,
Binh Duong, Vietnam

{vinh.nguyen, thanh.tran.cit20, doan.dang,
narayan.debnath}@eiu.edu.vn

Abstract. Vehicle detection and classification is an essential task in building an autonomous driving car. However, the accuracy of traditional vehicle detection systems is still not satisfied in real world due to unstable features from the original input images. This study introduces an efficient method to enhance the input feature more stable by considering the benefit of the machine learning approach to discover and extract robust features. The feature generated by the proposed deep learning model is then input to SVM to accurately identify objects. To evaluate the performance of the compared methods, we conducted comprehensive experiments on various datasets, such as KITTI, CCD, and HCI. For the KITTI dataset, the recall of our system, the LS-Support vector machine, the Linear support vector machine, the SVM-HOG, and the support vector machine are 71.4%, 64.3%, 53.8%, 67.8.0%, and 57.1%, respectively.

Keywords: Vehicle detection · Deep learning-based feature · Support vector machine · Hybrid learning

1 Introduction

Based on the news from WHO, many car accidents occur due to the driver carelessness [1]. Therefore, it is necessary to look for a solution to reduce vehicle accidents to save human life. Recently, research in vehicle detection and classification are concern by many research institutes and universities. Many methods and approaches have been proposed to increase the performance of vehicle detection and classification systems. The existing vehicle detection and classification systems, such as SVM [2], SVM-HOG [3], Linear SVM [4], and LS-SVM [5], are mostly trained and tested on the raw input images. If the input features are affected by noise, their performances also are decreased. Therefore, somehow, we need to find an approach to solve these limitations. This research studies a new approach by considering the advantages of the deep convolutional neural network to generate robust features from the original raw images.

The following section of this paper is organized as follows: we briefly review the existing vehicle detection and classification system in Sect. 2. From the weakness of the existing algorithms, we proposed the deep learning model for extracting robust features

to detect vehicles in Sect. 3. To evaluate the effectiveness of the proposed, various testing experiments were conducted on different datasets presented in Sect. 4. Finally, we point out the limitations of our proposed method and plan for future work in Sect. 5.

2 Related Works

Research in obstacle detection and classification algorithm has been extracted by many researchers. Without loss of generality, researchers have classified the existing obstacle detection and classification algorithm into two main classes: one stage-based method and two stage-based methods. In two stage-based methods, obstacle detection and classification algorithms [6, 7] often start with generating many proposal candidates on the input image. After that, the proposed candidates are then input into the detection and classification stage to accurately detect the object in the input image. In one stage-based method [8], obstacle detection and classification algorithms do not need to go through the first stage of the generation proposal candidate, the input images are passed directly to the detection and classification steps. Researchers found that the two-stage-based method achieves more accuracy than the one stage-based method because of the benefits of generating proposal candidates on the input images. However, the processing time of two-stage-based methods is much higher than one-stage-based methods. In general, we realized that both one-stage and two stages-based methods often use the raw input data, directly, to detect and classify objects. Therefore, somehow, the performance of existing obstacle detection and classification might decrease if the input image was affected by noise. This research studies a new approach by using the advantages of the deep convolutional neural network to generate robust features from the original raw images.

3 The Proposed Method

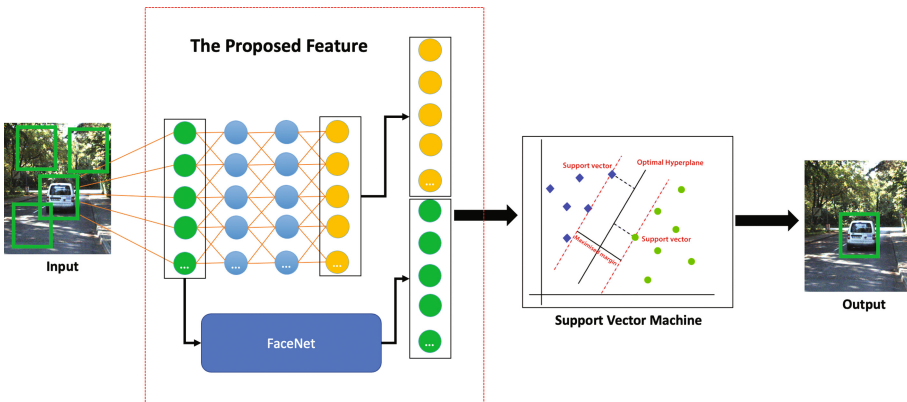


Fig. 1. Flow chart of the proposed system for vehicle detection

The proposed system includes two (Fig. 1): (1) the input image is inputted to the proposed deep CNN to extract robust features. (2) the robust feature is then inputted into the SVM [2] to detect and classify the vehicle.

Tradition machine learning algorithms use the raw image to train and detect objects. If the input size of the raw image is huge, the processing time of the existing methods also is increased. In addition, if the raw input data is not stable, the accuracy of the traditional obstacle detection and classification are significantly degraded. Therefore, we aim to design and develop an efficient deep learning model-based convolutional neural network to discover and extract features from the raw image.

Our proposed deep learning-based convolutional neural network is designed with 8 layers as described in Fig. 2. The first layer was designed by using the fixed input image size (128,128). The first hidden layer was designed with 1024 units in order to discover the abstract features from the 1st layer, followed by drop-out layer with a threshold of 0.5. The second hidden layer was designed with 512-units with the drop-out layer with a threshold of 0.5. The third hidden layer was designed with 256-units with the drop-out layer with a threshold of 0.5. The fourth hidden layer was designed with 128-units, with the drop-out layer with a threshold of 0.5. Finally, the output layer was designed with 2 units with the SoftMax classification. The total number of parameters of the proposed network is 17,467,522. Figure 3 describes our python code of our method.

Layer (type)	Output Shape	Param #
dense_36 (Dense)	(None, 1024)	16778240
dropout_32 (Dropout)	(None, 1024)	0
dense_37 (Dense)	(None, 512)	524800
dropout_33 (Dropout)	(None, 512)	0
dense_38 (Dense)	(None, 256)	131328
dropout_34 (Dropout)	(None, 256)	0
dense_39 (Dense)	(None, 128)	32896
dropout_35 (Dropout)	(None, 128)	0
dense_40 (Dense)	(None, 2)	258
=====		
Total params: 17,467,522		
Trainable params: 17,467,522		
Non-trainable params: 0		

Fig. 2. System configuration of the proposed method for feature extraction

```

model = Sequential()
model.add(Dense(1024, input_shape=(409600,), activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(512, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(256, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(2, activation='softmax'))

```

Fig. 3. Python code of the proposed deep learning model for feature extraction

After finishing training, the proposed system, we remove the last layer for future training. Thus, the proposed method successfully reduces the input feature size from 16384 units to 128 units. Recently, Schroff et al. introduced a novel method for extracting robust features in face recognition systems, named as FaceNet [10]. Therefore, this research also uses FaceNet to learn and extract 128 embedded features from 16384 units of the input image. Finally, the 128 features from the proposed method and 128 embedded features from the FaceNet are concatenated to create 256 robust features as shown in Fig. 2. The next step is to train the proposed 256 feature by using the support vector machine [11] to detect and classify the vehicle from the input image.

4 Experimental Results

4.1 Datasets, Evaluation Methods

There are various datasets were used to evaluate the accuracy of existing obstacle detection methods. This research conducted experiments on three datasets, the KITTI dataset [9], the HCI Dataset [12], and the CCD dataset [7].

To analyze the results of the proposed method and related research, we conduct experiments and compare them to various object detection methods, such as SVM [2], SVM-HOG [3], Linear SVM [4], and LS-SVM [5].

To measure the accuracy of the object detection method, precision ∂ and recall metric β , and F1 score are calculated as follows:

$$\partial = \frac{\delta}{\sigma + \delta}$$

$$\beta = \frac{\sigma}{\nu + \sigma}$$

$$F1 = 2 \frac{\partial * \beta}{\partial + \beta}$$

where δ is a result of true positive detection. σ is a result of number of false positive detection results. ν is a number of false negative detection results.

4.2 Results and Discussion

To evaluate the precision (PR), recall (RE), and f1 score of our system and compare methods, we set up comprehensive experiments on various datasets, such as KITTI, CCD, and HCI.

First, we evaluated the performance of our system on the KITTI datasets. We selected 200 images from the KITTI dataset in order to evaluate the precision, recall, and f1 score of compared methods. Figure 4 describes the results of our method, the SVM, the SVM-HOG, the Linear SVM, and LS-SVM on the KITTI datasets. As the visualization of the results, the proposed method with the SVM model got the best performance in comparison to existing methods. These results proved that our proposed feature is robust to enhance the performance of the vehicle detection system. We set up experiments with 200 images from the KITTI dataset. The comprehensive experiments are shown in Table 1. Analyzing the results from Table I, we found that the proposed method still gets the best performance. The precision of our method, the LS-Support vector machine, the Linear Support vector machine, the SVM-HOG, and the Support vector machine are 90.9%, 90%, 77.7%, 90.0%, and 88.9%, respectively. The recall of the proposed method, the LS-Support vector machine, the Linear Support vector machine, the SVM-HOG, and the Support vector machine are 71.4%, 64.3%, 53.8%, 67.8.0%, and 57.1%, respectively. The f1 score of the proposed method, the LS-Support vector machine, the Linear Support vector machine, the SVM-HOG, and the Support vector machine are 80.0%, 75.0%, 63.6%, 77.5%, and 69.6%, respectively.

Second, we analyzed the performance of our system on the CCD datasets. The KITTI dataset was captured under normal driving conditions. Thus, to verify the performance of our method on more challenging datasets, we select the CCD datasets. The CCD dataset provides images captured under adverse driving conditions. We also selected 200 images from the CCD dataset in order to evaluate the PR, RE, and F1 of compared

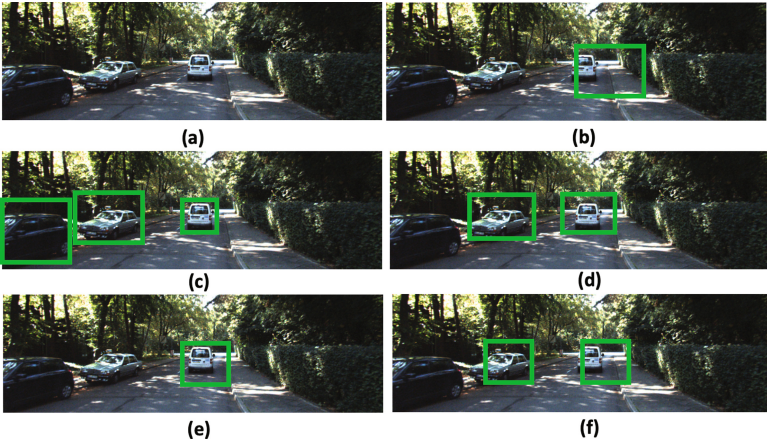


Fig. 4. Experimental results of our system and compared method on KITTI dataset. (b), (c), (d), (e), and (f) are visualization results of the SVM, our method, Linear SVM, LS-SVM, and SVM-HOG on the input image (a).

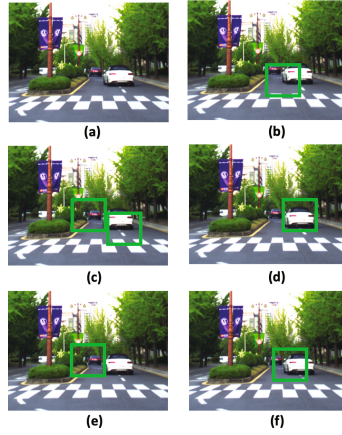


Fig. 5. Experimental results of our system and compared method on CCD dataset. (b), (c), (d), (e), and (f) are performance results of SVM, our method, Linear SVM, LS-SVM, and SVM-HOG on the input image (a).

methods. Figure 5 describes the performance of our system, the SVM, the SVM-HOG, the Linear SVM, and LS-SVM on the CCD datasets. As the visualization of the results, the proposed deep learning-based SVM model got the best performance in comparison to existing methods. These experimental results shown that our deep learning feature is robust to improve the accuracy of the vehicle detection system. Rather than verifying the accuracy of our system on a single image (as in Fig. 5). We set up experiments with 200 images from the CCD dataset. Table 2 shown the comprehensive experiments. Analyzing the results from Table I, we found that the proposed method still gets the best performance. The precision of the proposed method, the LS-SVM, the Linear SVM, the SVM-HOG, and the SVM are 89.6%, 88.2%, 79.4%, 88.9%, and 72.2%, respectively. The recall of the proposed method, the LS-SVM, Linear SVM, SVM-HOG, and SVM are 62.10%, 50.0%, 59.2%, 53.3%, and 52.0%, respectively. The f1 score of our method, the LS-SVM, the Linear SVM, the SVM-HOG, and SVM are 73.4%, 63.8%, 67.8%, 66.7%, and 60.5%, respectively.

Finally, we verified the performance of our system on the HCI datasets. The HCI dataset provides images captured under hostile driving conditions. Similar to the experiments on the KITTI and CCD datasets. We also selected 200 images from the HCI dataset in order to evaluate the PR, RE, and F1of compared methods. Figure 6 describes the experiments of our method, SVM, SVM-HOG, Linear SVM, and LS-SVM on the HCI datasets. As the visualization of the results, the proposed deep learning with the SVM model got the best performance in comparison to existing methods. These experimental results stated that the proposed deep learning feature is robust to improve the performance of the vehicle detection system. Rather than verifying the result of our proposed system on a single image (as in Fig. 6). We set up experiments with 200 images from the HCI dataset. The comprehensive experiments are shown in Table 3. Analyzing the results from Table I, we found that the proposed method still gets the best performance. The precision of the our system, LS-SVM, Linear SVM, SVM-HOG, and the SVM are

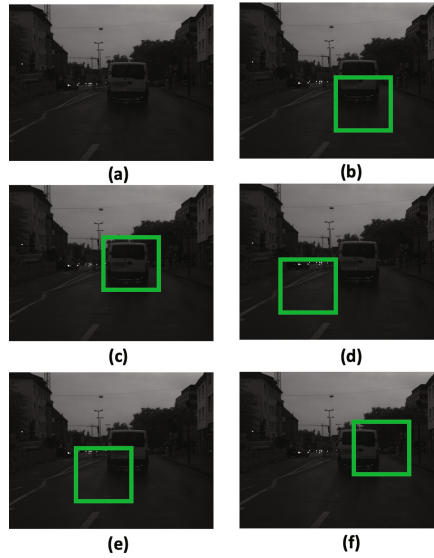


Fig. 6. Experimental results of our system and compared method on the HCI dataset. (b), (c), (d), (e), and (f) are performance of SVM, our method, Linear SVM, LS-SVM, and SVM-HOG on the input image (a).

88.1%, 71.8%, 76.7%, 87.5%, and 68.7%, respectively. The RE of the proposed method, LS-SVM, Linear SVM, SVM-HOG, and the SVM are 52.9%, 44.9%, 50.8%, 50.0%, and 44.0%, respectively. The f1 score of the proposed method, the LS-SVM, Linear SVM, the SVM-HOG, and SVM are 66.1%, 55.2%, 61.1%, 63.6%, and 53.7%, respectively.

Table 1. Results under the KITTI

	KITTI		
	Precision (%)	Recall (%)	
SVM	88.9	57.1	69.6
SVM-HOG	90.0	67.8	77.5
Linear SVM	77.7	53.8	63.6
LS-SVM	90.0	64.3	75.0
The proposed	90.9	71.4	80.0

Table 2. Results under the CCD datasets

	CCD		F1(%)
	Precision (%)	Recall (%)	
SVM	72.2	52.0	60.5
SVM-HOG	88.9	53.3	66.7
Linear SVM	79.4	59.2	67.8
LS-SVM	88.2	50.0	63.8
The proposed	89.6	62.10	73.4

Table 3. Experimental results using the HCI datasets

	HCI		F1(%)
	Precision (%)	Recall (%)	
SVM	68.7	44.0	53.7
SVM-HOG	87.5	50.0	63.6
Linear SVM	76.7	50.8	61.1
LS-SVM	71.8	44.9	55.2
The proposed	88.1	52.9	66.1

5 Conclusion

This research introduces the benefits of the proposed deep learning approach for improving the performance of existing vehicle detection systems. By conducting experiments under various datasets, including CCD, KITTI, and HCI, we found that the precision, recall, and f1 scores of the proposed method are better than compared method under various experimental conditions. However, there are still several limitations in the proposed method, such as increasing the processing time due to the pre-processing stage to generate robust features of our model. The lightweight approach should be considered to reduce the complexity of the proposed network.

Acknowledgements. The authors would like to express gratitude to Eastern International University (EIU), Binh Duong, Vietnam, for funding our research.

References

1. Road traffic injuries, WHO (2020). <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries#:~:text=Approximately%201.3%20million%20people%20die,result%20of%20road%20traffic%20crashes>

2. Li, H., Zhang, Z., Chen, W.: Detection for vehicle's overlap based on support vector machine. In: International Conference on Information Management, Innovation Management and Industrial Engineering, pp. 410–412 (2009)
3. Nan, M., Li, C., JianCheng, H., QiuNa, S., JiaHong, L., GuoPing, Z.: Pedestrian detection based on HOG Features and SVM realizes vehicle-human-environment interaction, In: International Conference on Computational Intelligence and Security (CIS), pp. 287–29 (2019)
4. Bougharriou, S., Hamdaoui, F., Mtibaa, A.: Linear SVM classifier based HOG car detection. In: International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), pp. 241–245 (2017)
5. Guangying, G., Cunwei, T., Minggong, W.: On the study of moving objects detection and pattern recognition using LS-SVM. World Congress on Intelligent Control and Automation, pp. 2486–2490 (2008)
6. Nguyen, D.V., Tran, T.D., Byun, J.Y., Jeon, J.W.: Real-time vehicle detection using an effective region proposal-based depth and 3-channel pattern. *IEEE Trans. Intell. Transp. Syst.* **20**(10), 3634–3646 (2019)
7. Nguyen, D.V., Nguyen, V.H., Tran, T.D., Jeon, J.W.: Learning framework for robust obstacle detection, recognition, and tracking. *IEEE Trans. Intell. Transp. Syst.* **18**(6), 1633–1646 (2017)
8. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
9. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The KITTI vision benchmark suite. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354–3361 (2012)
10. Schroff, F., Kalenichenko, D. and Philbin, J.: “FaceNet: a unified embedding for face recognition and clustering. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815–823 (2015)
11. Zehang, S., Bebis, G., Miller, R.: On-road vehicle detection using Gabor filters and support vector machines. In: International Conference on Digital Signal Processing Proceedings, vol. 2, pp. 1019–1022 (2002)
12. Meister, S.: Outdoor stereo camera system for the generation of realworld benchmark data sets. *Opt. Eng.* **51**(2) (2012)



Arabic Vowels Recognition Using Envelope's Energy and Artificial Neural Network

Nesrine Abajaddi¹ (✉) , Youssef Elfahm¹ , Badia Mounir² ,
and Abdelmajid Farchi¹

¹ IMII Laboratory, Faculty of Sciences and Technics, Settat, Morocco
nesrineabajaddi@gmail.com

² LAPSSII Laboratory, High School of Technology, Safi, Morocco

Abstract. In recent years, speech recognition and enhancement have grown in importance. Similarly, the modulation domain has gained attention in speech applications because that provides a more compact representation. Motivated by the development in the modulation domain and speech recognition. This paper aims to identify Arabic vowels using the energy contained in the envelopes based on the artificial neural network. In this work, the network was tested with a single hidden layer. As a result of the comparison with four neural network architectures, the results of the proposed network architecture have been analyzed and discussed. The proposed approach based on the energy of the envelope shows reliable results and works well with the artificial neural network.

Keywords: Arabic vowels · Artificial neural network · Spectral
Center-of-Gravity

1 Introduction

Vowels are the basic speech units present in all human languages. All vowels are voiced sounds because they are produced with the vocal folds vibrating. Several languages, notably English, Arabic, German, and Chinese...etc., are analyzed the vowels [1, 2]. The Arabic language is a semitic ancient language, and there are several languages derived from the Arabic language. They have six vowels are /a, i, u, a :, i :, u : /, characterized by the duration of their sound as short and long. The short vowels are /a/, /i/, and /u/, while the long vowels are /a:/, /i:/ and /u:/ which are the basis in almost all Arabic dialects [3–5] and 60% to 70% of Arabic speech consists of vowels [6].

Many works have been published in Arabic, have an interesting owing to the reliability and precision of the outcomes. The research served the Arabic Text to Speech system. Paddock's study [7], is one of the oldest studies that demonstrate an interest to examine Arabic vowels, exploits the vowel of the Russian and Egyptian participants. Alghamdi performed a study in which he evaluated whether the six vowels are at the same phonetic in different Arabic dialects [8], assuming that they are at the same phonological level. His analysis indicated that the phonetic implementation of the MSA vowel system varies depending on the spoken dialect. In [9], a comparison between long and short

vowels was performed for three languages, standard Arabic, Japanese, and Thai. The results stated that the duration of long vowels is twice that of short vowels. In another study by Seddiq and Alotaibi [10], are used the formants of the six Arabic vowels for vowel identification and characterization. Recent research [11] exploited the energy in the modulators to identify the vowels /a/, /i/, and /u/, they found that the modulator's energy can classify the Arabic vowels. The artificial neural network is used in much research and demonstrates its efficiency. Therefore, the work in [12], proposed the artificial neural network to recognize the short and long vowels using the formants as the inputs of the neural network.

This work aims to classify the short Arabic vowels using a simple feature, energy in the envelope, as an input of the artificial neural network which can improve the accuracy of classification of the Arabic vowels. What distinguishes our work from other works is the data that we collected ourselves in the laboratory, unlike other research which used ready data, simple and clean, recorded in noise-free areas. Besides, our work achieves a good recognition rate compared to other existing methods [12].

This paper is organized as follows, the proposed method which presents the extracted features and the used methods. Then, the results and discussion to analyze the outcomes of our proposed algorithm. Finally, the conclusion and future work.

2 Methods

In this section, we present the research methods adopted which provides a broader context for our work.

2.1 Corpus

The Arabic vowels corpus used in this work, is constructed by requesting twenty men to pronounce the vowels /a/, /i/, and /u/, each record contains five vowels for example /aaaaa/, /iiiiii/ and /uuuuu/, are repeated ten times for a person. These recordings are recorded in isolated room in the laboratory using "Praat". The vocal sound processing software, at 22050 Hz as the sampling frequency.

2.2 Proposed Method

In this paper, a new schematic approach is proposed to classify the Arabic vowels, based on the energy of the envelopes. In the first stage, the speech signal is divided into subband signals using the analysis filter bank (Fig. 1). Then, the subband signals are demodulated using the coherent demodulation method, to obtain the energy of the envelopes and exploit its characteristics at vowel recognition.

Speech can be expressed by sum of amplitude-modulated (AM) signals in frequency subbands covering the signal bandwidth because it is considered as modulated signal. The subband signal is contain a pair, envelope (modulator) and the fine structure (carrier). The full-band speech in discrete-time is defined by

$$x(n) = \sum_{k=1}^N s_k(n) = \sum_{k=1}^N e_k(n) \circ c_k(n) \quad (1)$$

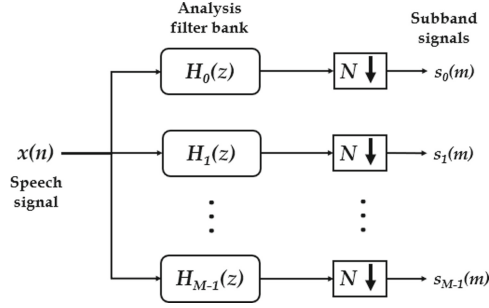


Fig. 1. The analysis filter bank

where $x(n)$ is the speech signal, $e_k(n)$ and $c_k(n)$ describe the k^{th} envelope and the fine structure of the subband $s_k(n)$, respectively, and N is the subbands number. To extract the envelope/modulator from the subband signal, the non-coherent demodulation is used, the envelope does not exceed in the subband limits in comparison with the incoherent method, which enables to define the energy characteristics of the Arabic vowels.

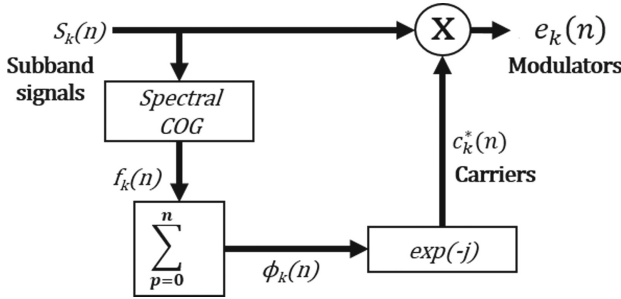


Fig. 2. The bloc diagram of coherent demodulation approach based on spectral Center-Of-Gravity

The envelope and the fine structure are respectively expressed in Eq. 2 and 3. Coherent subband demodulation is used with the Spectral Center-of-Gravity (COG) method, presented in Fig. 2, to track the evolution of spectral concentration within a subband, as an estimate of the instantaneous frequency $f_k(n)$. The carrier phase is defined in Eq. 4.

$$e_k(n) = |s_k(n)| \quad (2)$$

$$c_k(n) = \exp(j\angle s_k(n)) = \exp(j\phi_k(n)) \quad (3)$$

$$\phi_k(n) = \sum_{p=0}^n f_k(p) \quad (4)$$

The frequency bands used in this work, are five: B1(100–400 Hz), B2(400–800 Hz), B3(800–2000 Hz), B4(2000–3500 Hz) and B5(3500–5000 Hz). The examination of the average energy and the temporal variations of the envelopes shows that five bands allow

to identify the Arabic vowels. The envelope in the subband was framing by hamming window (256 points with an overlap of 50%). The percentage energy of a frame of the envelope written as follows

$$P_{ET} = \frac{E(b,j)}{E_T(j)} * 100 \quad (5)$$

with $E(b, j) = \sum_{k=0}^{i-1} |m(b, k)|^2$ is the frame energy, $E_T(j) = \sum_{n=0}^b E(j, n)$ is the total energy of all envelopes for the same frame. Where j is the frame number, i is the frame length and b is the number of the envelopes or bands.

2.3 Artificial Neural Network

Neural networks, also defined as artificial neural networks (ANNs), are a subfield of machine learning that serve as the base of deep learning. ANNs are constituted of node layers that include an input layer, hidden layers, and then an output layer. Training data is used by neural networks to train and enhance their accuracy. Moreover, once these learning algorithms have been good for accuracy. They become reliable tools in computer science and artificial intelligence, helping us to identify and organize data at high speed when compared with human manual verification. Google's search algorithm is one of the most well-known neural networks. As a result, the ANN can execute complicated tasks such as identification and classification, because of its excellent learning characteristics. In general, there is no defined method for calculating neuron numbers in the hidden layer. However, the hidden layer must have the same size as the input layer or 75% of its size as indicated in [13].

To recognize the three vowels /a/, /u/, and /i/, the network has five input layers that are controlled by the energy of envelopes (five bands). Many transfer functions are used in the Neural Network. The sigmoid transfer function is applied in the hidden layer. The ANN is evaluated with a single hidden layer with four neurons (75% of the input size).

3 Results and Discussion

To ensure that our experiments were performed correctly, the corpus of the speech was divided into three datasets: 70% of the data for training database, 25% of the total database for testing, and 5% for validation. The major issue is that ANNs cannot explain their predictions or the mechanisms during network training which prevents interpretation, but it is still in development.

The confusion percentages of the vowels are 0.67% and 2.44% (Table 1), between the vowels /u/ and /i/ or /a/. For /u/ with /a/, the confusion is derived from the energy distribution of /u/ in the first two envelopes (B1 and B2) as mentioned in [11]. And for the vowel /i/ some speakers can generate a higher value of the second formant F2 (>3500 Hz), the energy in the fourth envelope decreases which is induced the confusion between /u/ and /i/ [9].

Table 1. The confusion matrix

Vowels	/a/	/i/	/u/
/a/	100	0	0
/i/	0	99.33	0.67
/u/	2.44	0	97.56

According to the analysis of the results, the most challenging vowels in classification are those with a similarity to /u/. The best vowel to categorize is /a/ because it does not have the most similarity to the other vowels. The performance of the proposed method is good and accurate. The energy of the envelopes works well with the artificial neural network. The obtained performance shows that don't need to add other hidden layers or change the number of neurons, because of the satisfactory results.

Table 2 shows a comparison of the developed method with the method defined in [12], they used the formants as inputs of ANN. The performance of the proposed method is greater than results of Aloqayli [12], for all the vowels /a/, /i/, and /u/. Therefore, it can be concluded that the proposed method is simple in its implementation, whose recognition performance is very interesting but needs training data for better accuracy.

Table 2. Recognition rate of the three vowels /a/, /u/ and /i/, comparison of the proposed method with the results presented in [12]

Vowels	Proposed method, one Hidden Layers with 4 neurons	ANN [12], single Hidden Layers with 16 neurons	ANN [12], single Hidden Layers with 30 neurons	ANN [12], two Hidden Layers with [5–7] neurons	ANN [12], two Hidden Layers with [8–8] neurons
/a/	100	100	88.89	83.33	83.33
/i/	99.33	83.33	83.33	72.22	77.78
/u/	97.56	72.22	83.33	83.33	100

4 Conclusion

In this work, a recognition vowel system using an artificial neural network and the energy contained in the envelopes is proposed in order to identify the Arabic vowels. The envelopes are calculated using coherent demodulation based on the spectral Center-of-Gravity and the neural network with a single hidden layer. According to our results, the system had a reliable performance for all three vowels /a/, /i/, and /u/.

The goal for future work is to add more vowel features and examine more complex ANN systems. More analytical work in vowels analysis or consonants will improve the literature and provide a good base for researchers and developers in many fields of research as speech recognition, speech detection, and speech intelligibility.

References

1. Szalay, T., Benders, T., Cox, F., Proctor, M.: Perceptual vowel contrast reduction in Australian English /l/-final rimes. *Lab. Phonol.* **12** (2021)
2. Barkana, B.D., Patel, A.: Analysis of vowel production in Mandarin/Hindi/American-accented English for accent recognition systems. *Appl. Acoust.* **162** (2020)
3. Newman, D.L.: The phonetic status of Arabic within the world's languages: the uniqueness of the lu‘At Al-d÷AAd Daniel. *Antwerp Pap. Linguist.* **100** (2002)
4. Alotaibi, Y.A., Husain, A.: Formant based analysis of spoken arabic vowels. In: Fierrez, J., Ortega-Garcia, J., Esposito, A., Drygajlo, A., Faundez-Zanuy, M. (eds.) *BioID 2009. LNCS*, vol. 5707, pp. 162–169. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04391-8_21
5. Alotaibi, Y.A., Hussain, A.: Speech recognition system and formant based analysis of spoken arabic vowels. In: Lee, Y.-H., Kim, T.-H., Fang, W.-C., Ślęzak, D. (eds.) *FGIT 2009. LNCS*, vol. 5899, pp. 50–60. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10509-8_7
6. Nabil, A., Hesham, M.: Formant distortion after codecs for Arabic. In: *Final Program and Abstract Book - 4th International Symposium on Communications, Control, and Signal Processing, ISCCSP 2010* (2010). <https://doi.org/10.1109/ISCCSP.2010.5463385>
7. Paddock, H.J.: The major pitch features of vocalic quality. *Lingua* **25** (1970)
8. Alghamdi, M.M.: A spectrographic analysis of arabic vowels: a cross-dialect study. *J. King Saud Univ.* **10** (1998)
9. Tsukada, K. An acoustic comparison of vowel length contrasts in standard Arabic, Japanese and Thai. in *2009 International Conference on Asian Language Processing: Recent Advances in Asian Language Processing, IALP 2009* (2009). doi:<https://doi.org/10.1109/IALP.2009.25>
10. Seddiq, Y.M., Alotaibi, Y.A.: Formant-based analysis of vowels in Modern Standard Arabic Preliminary results. In: *2012 11th International Conference on Information Science, Signal Processing and their Applications, ISSPA 2012* (2012). <https://doi.org/10.1109/ISSPA.2012.6310641>
11. Abajaddi, N. et al.: Efficiency of the energy contained in modulators in the Arabic vowels recognition. *Int. J. Electr. Comput. Eng.* **11**, (2021)
12. Aloqayli, F.M., Alotaibi, Y.A.: Spoken arabic vowel recognition using ANN. In: *Proceedings - UKSim-AMSS 11th European Modelling Symposium on Computer Modelling and Simulation, EMS 2017* (2017). <https://doi.org/10.1109/EMS.2017.24>
13. Venugopal, V., Baets, W.: Neural networks and statistical techniques in marketing research: a conceptual comparison. *Mark. Intell. Plan.* **12** (1994). <https://doi.org/10.1108/02634509410065555>



Soil Nutrient Prediction Model in Hybrid Farming Using Rule-Based Regressor

M. Krishnaveni¹ , Paa. Raajeswari² , P. Subashini¹ , V. Narmadha¹ ,
and P. Ramya²

¹ Department of Computer Science, Centre for Machine Learning and Intelligence,
Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India
narmadha_cs@avinuty.ac.in

² Department of Food Science and Nutrition, Avinashilingam Institute for Home Science and
Higher Education for Women, Coimbatore, India

Abstract. Agriculture is the supreme trade of India. Inefficiency and imprecise input controls such as nutrients of the soil, water, and usage of hazardous manure have caused devastating consequences to the biosphere. Various plant species are generated daily, however, they are deficient in all the necessary nutrients compared to the crop that is grown organically. To overcome this situation, an integrative hybrid approach is proposed in this paper, to combine both precision farming and organic farming which involves growing and fostering crops without the use of non-natural fertilizers and pesticides to elevate and enhance the quality and quantity of the crops. This paper proposes a machine learning (ML) model to predict nutritional values in Ballarat (Centella Asiatica) in both conventional farming and pro-biotic farming. This Hybrid system predicts the nutrient values such as nitrogen, phosphorus, and potassium with the supplement of banana peel powder prediction accuracy, mean absolute Error (MAE), Root mean squared error (RMSE), and R^2 for potassium evaluate the performance of the model. The results reveal that random forest regressor performs well in probiotic farming with 91% and RMSE is 1.7475 and MAE is 0.6361 than decision tree regressor.

Keywords: Organic farming · Precision farming · Pro-Biotic farming · Soil sensor (ECa) · IoT-based technologies

1 Introduction

Agriculture is a popular vintage of mankind. The fundamental aim of agriculture is to uplift growth by improving the soil, water, and other nutrient composition. In spite of the worldwide rapid growth of industrialization and urbanization, nearly one-half of the working population is still engrossed in agriculture [5]. In India, there are around 215.6 million acres (82.6 million hectares) of farmland used for agriculture, 48.92 lacks which are in Tamil Nadu. The UN Food and Agriculture Organization depicts that “The world has to produce 70% more food in 2050 than in 2006” [1]. The agricultural methods followed all over the globe are subordinate to the widespread use of insecticides and

fertilizers that incorporate synthetic formulations will therefore improve the crop quality and help feed the world's population [2]. On the other hand, it is impossible to ignore the numerous negative effects of chemical fertilizers and pesticides. Fertilizers have the ability to stay in the soil for a longer period of time and have negative drastic on a variety of biotic and abiotic elements, including the soil, the environment, and even human health [3]. Soil standards are often compromised by these man-made pesticides, leading to further environmental degradation. Agriculture is particularly vulnerable to devastating natural hazards such as insect damage and adverse weather conditions that affect crop yields. Financial Express states that about 16–20% of the total crops produced in India were wasted each year [4]. Analyses of suitable environmental conditions can improve yields and reduce damage and crop loss. Soil fertility management is the most important factor in maintaining good crop yields. The microelement or microbes is influential production of the plant life cycle. Lack of micronutrients in the soil results in abnormal growth of the plants which may lead to crop failure [6]. However, harvest losses can be minimized, and yields are increased by integrating with any kind of combinational farming method. Precision agriculture is the use of innovative technologies and principles to manage the geographic and terrestrial variability associated with all aspects of agricultural production, improve crop performance, increase crop productivity, to improve the quality of the environment [7, 9]. IoT-based precision agriculture improves livestock production by predicting fertility patterns and diagnosing eating disorders, and cattle behavior based on ML models [12]. Plant probiotic microorganisms (PPM), also known as biofertilizers, are useful microbes that suggest a favourable alternative method that reduces health problems and environmental degradation. The use of plant probiotics serves as an alternative soil fertilizer which has been the focus of several studies. The use of probiotic in farming upgrade nutrient values, sustain a suitable environment for field management, and creates no unfavorable effects [10]. From the different research work carried out so far, it is inferred that the conventional farming method is specific to either precision farming or organic farming. Integration of both precision farming and organic farming is proficient in sustaining higher crop productivity and enhancing soil quality on an unceasing basis. Therefore, this research work is motivated to compare and contrast the soil nutrient values i.e., Potassium(K) in two different types of soil in conventional farming and pro-biotic farming in AI IoT-enabled Soil Nutri farm situated in the Centre for Machine Learning and Intelligence in our university premises. With the advancement of sensor-based technologies along with organic farming monitors the field 24*7 and updates the values in the cloud server. Therefore, proper monitoring of soil conductivity, soil moisture, and soil temperature is predicted to enrich the nutrient values such as Nitrogen, Phosphorous, and Potassium of the crop to increase productivity.

The rest of the article is formulated as follows, Sect. 2 Literature study, Sect. 3 Data Acquisition, and Experimental Setup depict a detailed description of the data collection method. Section 4 Methodology, Sect. 5 describes about performance metrics, Sect. 6 denotes, Results and discussion and finally Sect. 7 includes conclusion.

2 Literature Study

The applications of advanced technology in precision farming have rapidly increased over a decade. The promising research works on the technological evolution of farming are summarized below.

Monteiro et al. [8] The control system is governed by Arduino and sensors such as the DHT11 sensor, soil moisture sensor (REES52), sound module(buzzer), and PIR sensor (HC-SR501), which provide the temperature, humidity, and moisture are used. Precision agriculture proves to be an efficient method to manage the resources such as crop yields, livestock, seeding, and fertilizer.

Reddy et al. [11] The paper proposed a two-feasibility study on IoT-based solutions for automated irrigation and animal monitoring. Each component comprises a temperature sensor (DHT11), Soil moisture sensor (REES52), and solar power cells with rechargeable batteries with a ZigBee module. A web interface is used for visualizing the irrigation schedule. It is also used to set threshold values on the web interface to automate the scheduling. A SQL database is used for storing the data.

Heiniger, et al. [13] conducted an experimental setup is done in 15 different field sites with 12 different soil series in three regions of North Carolina. Nutrient value and attributes of the soil are compared with electrical conductivity with the help of Correlation analysis and PC-stepwise regression analysis. The result states that few convincing relationships are found between electrical conductivity and nutrient concentrations.

The above reviews showed that the usage of precision farming is convenient with refinement over the traditional method. Soil electrical conductivity (EC_a) plays a major role in soil health. It is an indirect indicator of nutrient concentration in the soil. An experimental study was initiated to demonstrate whether (EC_a) could predict the NPK in two different types of soil.

3 Data Acquisition

3.1 Preparation of the Field

Prior to the seeding, the selected field is AI-IoT Based Nutri Garden in front of CMLI at our university. The selected field is 20 feet in length and 10 feet in height. The soil is a mixture of sand and clay with a soil conductivity of $0.11 \text{ (dsm}^{-1}\text{)}$, pH value of the soil is 7.86. The soil test has been carried in Tamil Nadu Agricultural University, Coimbatore, India. The Micro and Macro nutrients are calculated and the results state that the soil has a higher potassium rate of 165 (mg/g) , mild phosphorous with a rate of 6.7 (mg/g) , and lower Nitrogen with a rate of 53 (mg/g) . The seeds were collected from the Tamil Nadu Agricultural University seed center, Coimbatore. The seeds like Horse gram, Fennel, Fenugreek, Coriander, Amarnath, and Vallarai seeds are sowed in the selected field. The field area is divided into two types of farming i.e., one is conventional Farming and the other one is probiotic farming. In conventional farming, six different Indian crops like Horse gram, Fennel, Coriander, Fenugreek, Amarnath, and Vallarai. In the probiotic farming method, the six different Indian crops along with banana peel powder which acts as a microbe are sown. When banana waste is reapplied to the plant, it will keep the soil moist [19]. Banana is one of the most popular fruits for their high nutritional content

and it also has a significant economic influence [20]. Furthermore, banana waste can be used as a natural fertilizer which alternative to synthetic fertilizers. With the inference of the above literature, it is found that banana peel powder boosts the soil nutrient level. The banana peel powder is prepared by us by the following two methods i) Preparation of raw materials ii) Preparation of Banana peel fertilizer.

3.1.1 Preparation of Raw Materials

Banana peels are available in all seasons. Banana peel wastes are collected from chips shop located in Madhampatti, Coimbatore, Tamil Nadu, India. Banana peels are considered garbage, smell bad, and contain many chemical elements or compounds that are obviously beneficial to plants. Banana peel powder fertilizer is rich in potassium and magnesium, both of which enhance stem and plant root growth and improve plant nutrient levels.

3.1.2 Preparation of Banana Peel Fertilizer

Organic fertilizer helps in boosting up agricultural productivity in terms of both quality and quantity. It reduces soil pollution and increases soil quality naturally [14]. Banana peels should be cut into little squares that are about an inch wide. It should be dried in sun for 5 days at 29.8 °C. Then, it grounded into fine powder form. This fine powder of banana peel is ready to use in agriculture farms to enhance the nutrient content of soil particularly nitrogen, phosphorus, and potassium [15]. Figure 1 depicts the preparation of banana peel fertilizer i.e., (Fig. 1). Figure 2 (a) represents the Sowing of seeds (Fig. 2).

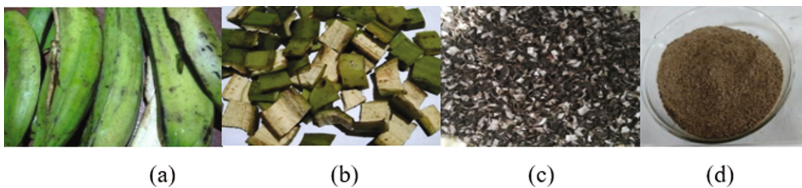


Fig. 1. Preparation of Banana Peel Fertilizer a) banana peel b) Cut into small pieces c) dried banana peel d) banana peel powder

3.2 Configuration and Deployment of IoT-Based Sensors in the Field

Precision farming is one kind of farming method employed to implement sensor-based farming. As an initial study, The LSE01 is a LoRaWAN Soil Moisture Sensor which is used for IoT in Agriculture is purchased and deployed. Sensor consists of two probes to calculate soil moisture, soil conductivity, soil temperature. Nitrogen, Phosphorous and Potassium values are observed in wet lab in Food Science and Nutrition laboratory. This sensor is used to measure soil moisture in saline-alkaline and loamy soils. The soil sensor uses the FDR method to calculate soil moisture by calibrating soil temperature and conductivity. It is also manufactured to identify soil types of industrial minerals.



Fig. 2. (a) Sowing of seeds (b) Discussion of the nutrient content

It detects soil moisture, soil temperature and soil conductivity and NPK. The collected data is sent to LoRaWAN IoT Server. LSE01 - soil conductivity sensor is deployed in both conventional farming and pro biotic farming. Figure 3 a) shows the deployed 5 sensors each of them in probiotic and conventional farming b) shows the dashboard.

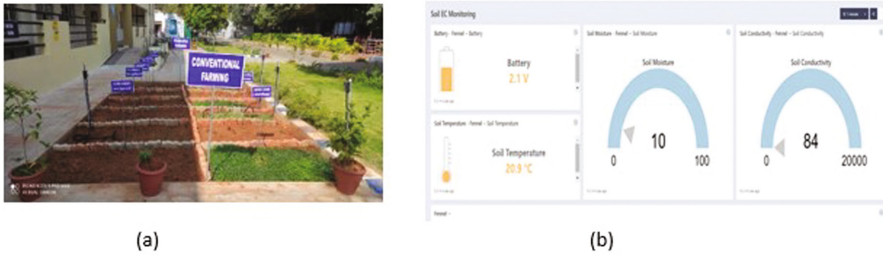


Fig. 3. a) Configuration of soil sensor in the farm b) Data monitoring of fennel seed cropping in the dashboard

3.3 Dashboard Creation

The dashboard has been created to visualize the data from anywhere in the world. The dashboard was developed in the Go programming language, which is an open-source programming language which is supported by google. It is a strong platform that supports Windows, Linux, and macOS. The Front end of the dashboard is developed in GO and the Back end is developed in React JS. The dashboard is an interactive toolkit, where we can monitor the parameters and status of the battery. The values can be extracted in three main file formats such as CSV, Excel and TSV file formats can be downloaded with the given user credentials. With the stipulated given time frequency, the datasheet can be downloaded in the desired file format.

4 Methodology

Innovative technological-based modelling, such as AI-IoT-based techniques can improvise and adopt vast amounts of data and acquire an inference for the future. From the above-proposed framework, the opportunities for deploying AI-IoT-enabled farming are demonstrated in Fig. 4.



Fig. 4. Framework for the proposed methodology

4.1 Data Collection

As an initial study, the data is collected from Vallarai (Centella Asiatica) in both conventional farming and pro-biotic farming. The time series data is extracted for the period of 31 days i.e., from 1.05.2022 to 31.05.2022, for every 30 min, the data will be collected and stored in the dashboard. The collected dataset is extracted as a comma-separated value file (CSV) format which is readily available in the dashboard. The data set consists of 1919 records with eight attributes Date Time, Soil Temperature, Soil Conductivity, Soil Moisture, Battery level from the field, and Nitrogen, Phosphorous, and Potassium from lab results. Two different datasets are collected from the field, one is probiotic farming, and another one is a traditional farming method. Table 1 depicts the sample dataset of conventional farming. Table 2 depicts the sample dataset of probiotic farming. It contains soil moisture, soil temperature, soil conductivity, N, P, and K values with respect to data and time (Table 1 and 2).

Table 1. Sample dataset for Conventional Farming

Date time	Soil moisture (%)	Soil temperature (°C)	Soil conductivity (dsm ⁻¹)	N	P	K
31-05-2022 21:24	12.42	26.92	150	50	33	80
31-05-2022 21:04	12.46	27.01	150	50	33	80
21-05-2022 23:44	20.32	26.31	276	54	32	82

Table 2. Sample dataset for Pro Biotic Farming

Date time	Soil moisture (%)	Soil temperature (Â°C)	Soil conductivity (dsm ⁻¹)	N	P	K
31-05-2022 21:24	15.29	27.4	192	58	49	157
31-05-2022 21:04	15.34	27.66	194	59	49	157
21-05-2022 23:44	21.25	29.19	297	59	46	156

4.2 Data Pre-processing

Data pre-processing is a technique for preparing data and creating models suitable for developing machine learning models. Data Cleansing is done by removing insignificant parameters such as the Battery (V) of the sensor, Date, and Time. The data is converted to a data frame using Panda’s package in python. With the help of the panda’s package insignificant field has been removed. The data format now consists of parameters like soil conductivity, soil temperature, soil moisture and Nitrogen, Phosphorous and Potassium.

4.3 Exploratory Data Analysis

Exploratory Data Analysis is done to evaluate and visualize the data. By choosing the most important features in the model. The correlation coefficient is used to measure the interrelation between two or more variables. It is also used to find the relationship between the two data and estimate how strongly dependent to each other. Correlation between two or more variables is calculated, where the total sample size is N.

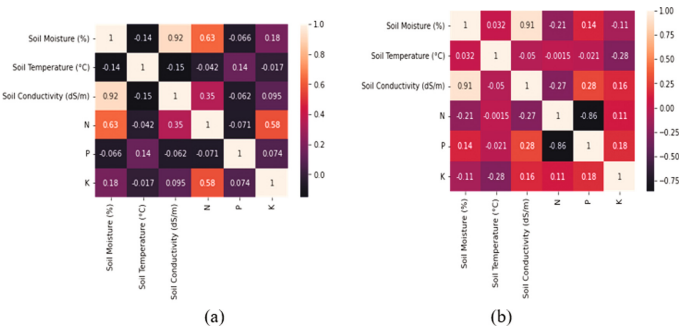


Fig. 5. Correlation between variables a) conventional farming b) probiotic farming

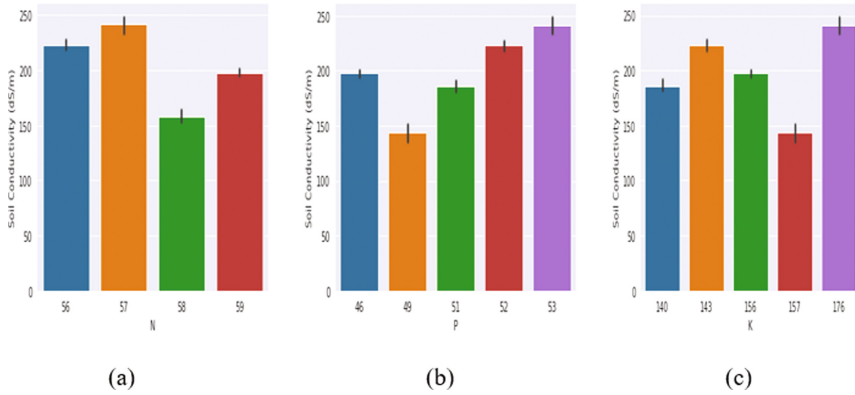


Fig. 6. Soil Conductivity Vs Nitrogen, Phosphorous, and Potassium in Pro Biotic Farming

Figure 5 represents the correlation between variables in two different dataset. To evaluate the relationship between soil conductivity and nutrient values such as NPK, a bar plot from the seaborn library is used to plot with the respect to each nutrient values and soil conductivity. From the above Fig. 6, it is inferred that nutrient values with respect to soil conductivity and NPK are found higher in pro-biotic framing when compared with conventional farming.

4.4 Rule-Based Learning

One of the data mining techniques extensively used in classifiers to classify or predict categorical class names. Classification algorithms have the potential to handle large amounts of information. It is used to hypothesize category names for classifying knowledge based on training data and class labels for classifying newly acquired data. Machine learning classification algorithms include multiple algorithms, and this paper focuses on the popular decision tree algorithm.

4.4.1 Decision Tree

A decision tree is widely used due to the following factors such as ease to use, clarity, and strong in detecting outliers [16]. The usage of a Decision Tree is much easier than numerical weights in neural networks. Decision Tree acts well in both discrete and continuous variables. Furthermore, in Data Mining DT is the most widely utilized a classification model. It maps the non-linear model quite well, when to compared to a linear model. It works on the divide and conqueror rule, where the hierarchical tree structure is formed.

A simple decision tree consists of target variable Y i.e. (0 or 1) and two continuous variables $\times 1$, $\times 2$. The components of a tree are nodes, branches, and importantly splitting, stopping, and pruning.

Algorithm

Training dataset: x_i ,

Label vector: $\hat{y}_i$

$threshold_c = 110 - 570 \text{ dsm}^{-1}$, $threshold_t = 20 - 30^\circ\text{C}$ $threshold_m = 10 - 45\%$

Load the dataset: Load the data sample, it consists of 1919 rows and 8 Columns.

1. **Split the dataset:**

Let the data at node m be represented by D_m with nm samples

The data is split as 80% for training and 20% for testing, for each node split $\omega = (f, threshold_m)$ where f is feature and limit $threshold_m$, split the data as

If x_i is empty then

Return leaf node:

End if

If $V_m^{left} \text{ soil Conductivity}(\theta) = \{(x, y)x_j \leq threshold_c\}$ **then**

Create node $V_m^{left} \text{ soil Conductivity}(\theta)$

if $V_m^{centre} \text{ soil temperature}(\theta) = \{(x, y)x_j \leq threshold_t\}$ **then**

Create node $V_m^{centre} \text{ soil temperature}(\theta)$

if $V_m^{left} \text{ soil moisture}(\theta) = \{(x, y)x_j \leq threshold_m\}$ **then**

Create a node $V_m^{right} \text{ soil moisture}(\theta)$

End If:

If $V_m^{left} \text{ soil Conductivity}(\theta) \leq 110 -$

$570 \text{ dsm}^{-1} \parallel V_m^{centre} \text{ soil temperature } 20 - 30^\circ\text{C} \parallel V_m^{right} \text{ soil moisture}(\theta) \leq 10 - 45\%$

Create a node **Nitrogen, Potassium, Phosphorous**

3. Compute Impurity function

The quality of a split of node m is can be computed using an impurity function or loss function H()

$$G(D_m, \theta) = \frac{n_m^{left}}{nm} H(V_m^{left} \text{ soil Conductivity}(\theta) + \frac{n_m^{centre}}{nm} H(V_m^{centre} \text{ soil temperature}(\theta) + \frac{n_m^{right}}{nm} H(V_m^{right} \text{ soil moisture}(\theta))$$

$$\theta^* = \operatorname{argmin}_{\theta} G(D_m, \theta)$$

For each node:

Calculate Standard deviation

$$SD(S) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

Where y_i is target values, $\bar{y} = \frac{\sum}{n}$ is mean value and n is the number of splits.

$$S(T, P) = \sum_{n \in x_i} X(c)S(c)$$

Where S is SD, x_i is dataset is Target variable potassium, P is Predictor variable

$$S(\text{potassium}, x_i) = X(\text{Soil Conductivity}) * S(\text{Soil Conductivity}) + X(\text{Soil Moisture}) * S(\text{Soil Moisture}) + X(\text{Soil Temperature}) * S(\text{Soil Temperature})$$

Return Outcome Potassium

Repeat until n=7

End

4.4.2 Random Forest

Random Forest is a grouping technique which is capable to perform both classification and regression. Random Forest outperforms technique called ensemble or bagging approach, which combines a one or more classifiers to solve a complex problem and improves the performance of the model. Random Forest [18]. The greater number of trees in Random Forest, greater the predictions. Advantages of Random Forest algorithm prevents the data from over fitting. The approach associates several randomly generated trees and sum their prediction by averaging. In addition to it, it is adaptable to many large-scale problems and flexible to different extemporary learning task. Random forest can be applied in various sectors like banking, healthcare to predict and classify the outcome. Random forest is also known as group of decision trees which is a flexible algorithm in this machine learning era. Bagging is common grouping strategy which groups the samples from the original dataset, build a predictor model from each of its sample and determine by averaging.

A Random Forest is a forecasting method, which consists of group of T randomized regression-based trees. For the i_{th} tree in the group, the predicted value at the query point a , is represented as $t_n(a; \theta_i, D_n)$ where $\theta_1 \dots, \theta_t$ are independent random variables, distributed as common random variable θ which is independent of Dataset D_n . The variable θ is used to resemble the training is set prior to the growth of individual trees and helps in selecting right path for splitting. In mathematical form, i_{th} tree will be estimated as

$$T_n(a; \theta_i, D_n) = \sum_{j \in D_n^*(\theta_i)} \frac{1_{a_i \in A_n(a; \theta_i, D_n) B_i}}{N_n(a; \theta_i, D_n)} \quad (1)$$

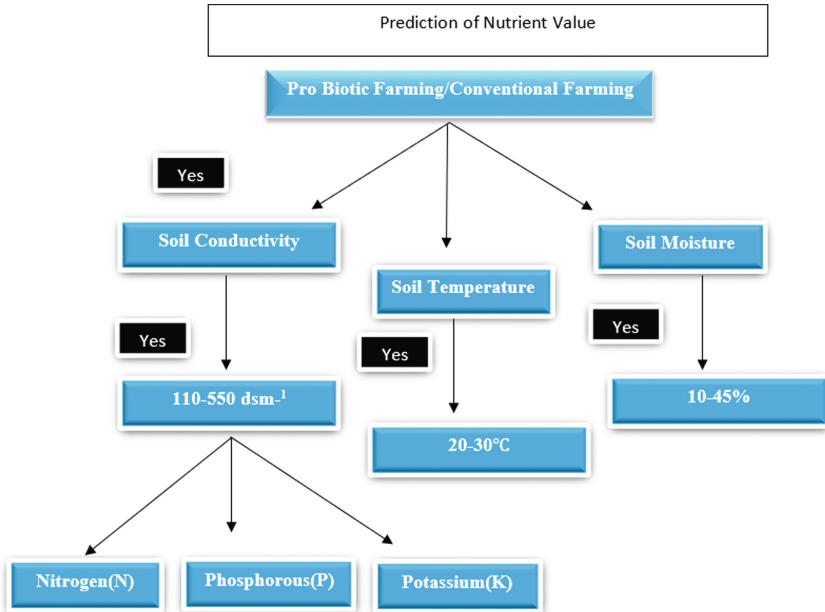


Fig. 7. The prediction of nutrient values

where $D_n^*(\theta_i)$ is the set of data selected for construction of the tree, $A_n(a; \theta_i, D_n)$ is the cell containing, $N_n(a; \theta_i, D_n)$ is the number of points which fall into $A_n(a; \theta_i, D_n)$. The trees are combined to form a finite forest (Fig. 7).

$$t_{T,n}(a; \theta_1, \dots, \theta_T, D_n) = \frac{1}{T} \sum_{i=1}^T T_n(a; \theta_i, D_n) \quad (2)$$

Algorithm

Inputs

Training dataset: $D = x_1, y_1 \dots x_n, y_n$ $threshold_c = 110 - 570 \text{ dsm}^{-1}$, $threshold_t = 20 - 30^\circ\text{C}$, $threshold_m = 10 - 45\%$

1. **Load the dataset:** Load the data sample, it consists of 1919 rows and 8 Columns.

2. **Split the dataset:**

Let the data at node m be represented by D_m with nm samples. Random Forest constructs many decision trees The data is split as 80% for training and 20% for testing, for each node split.

3. Sample S sets of n elements from D with replacements

Construct a Tree

If $V_m^{left} \text{ soil Conductivity}(\theta) = \{(x, y)x_j \leq threshold_c\}$ **then**

Create node $V_m^{left} \text{ soil Conductivity}(\theta)$

if $V_m^{centre} \text{ soil temperature}(\theta) = \{(x, y)x_j \leq threshold_t\}$ **then**

Create node $V_m^{centre} \text{ soil temperature}(\theta)$

if $V_m^{left} \text{ soil moisture}(\theta) = \{(x, y)x_j \leq threshold_m\}$ **then**

Create a node $V_m^{right} \text{ soil moisture}(\theta)$

End If:

If $V_m^{left} \text{ soil Conductivity}(\theta) \leq threshold_c \parallel V_m^{centre} \text{ soil temperature} \leq threshold \parallel V_m^{right} \text{ soil moisture}(\theta) \leq threshold_m$

Create a node **N, P, K**

For each $D_{i=1 \dots S}$

obtain a sequence of S outcome

$N(x), P(x), K(x) \dots N_t(x) P_t(x) K_t(x)$

For each tree:

Calculate Feature's Importance

$f_i = \sum_{j: \text{node } j \text{ splits on features } i} S_j C_j$,

where f_i is important, S_j number of sample reaching node j C_j , impurity of node j

Compute Final Feature importance

$$RF f_{i_i} = \frac{\sum_j norm f_{i_j}}{\sum_{j \in \text{all features}, k \in \text{all trees}} norm f_{j_k}}$$

End

5 Performance Metrics

Accuracy is a simple metric that measures the number of correct predictions over the total number of predictions.

$$\text{Accuracy} = \frac{\text{Number of Correct predictions}}{\text{Total Number of predictions}} \quad (3)$$

R-Squared (R^2) is a common statistical measure used in regression model which determines the proportion of variations in the dependent variable by the independent variable, also r-squared illustrates how good the data is fit in the regression model.

$$R^2 = 1 - \frac{\text{UnknownVariations}}{\text{TotalVariations}} \quad (4)$$

Root Mean square (RMSE) is the square root of variance of the predicted errors. RMSE is the popular method of identifying the error rate in the model.

$$\text{RMSE} = \sqrt{\sum (P_i - \frac{O_i}{n})^2} \quad (5)$$

where $\sum$ denotes Sum, P_i denotes predicted value in the dataset, O_i observed value in the data set. Mean Absolute the simplest error metric to evaluate the error rate of model [17]. It is effectively used in all data mining problems. It is approximately the sum of the average absolute differences between the predicted and actual values $\text{MAE} = \frac{1}{n} \sum |X - \hat{X}|$ where $\sum$ denotes Sum, X denotes predicted value in the dataset, value in the actual value set.

6 Results and Discussion

In this research work, as an initial study with Vallarai (*Centella asiatica*) is considered to perform the comparative analysis between conventional farming and hybrid farming. This paper provided an inference that, soil conductivity plays a major role in enriching the nutrients NPK of the soil. It is inferred that when soil conductivity is within threshold of 110–150 dsm⁻¹, temperature is between 20–30 °C, soil moisture is within 10–45% the values of nitrogen, potassium and phosphorous is found higher than usual. The carried work uses regression analysis to predict the nutrient value (NPK) with the respect to soil conductivity. The Decision Tree Regressor and Random Forest Regressor has been implemented to compare between pro biotic farming and conventional farming and their limitations were conducted. In Decision tree and Random Forest soil conductivity is higher in pro biotic farming compared with conventional farming. Results proved that the banana peel powder acts as an immune booster for the soil which enhanced the potassium level to the greater extent. Performance metrics such as accuracy, RMSE and MAE is calculated for Decision Tree. R^2 , RMSE and MAE is calculated for Random Forest. In conventional farming the soil has less conductivity than probiotic farming. Accuracy rate for Decision Tree in Pro Biotic farming is 85% whereas, conventional farming it is 69%. R square rate for Pro Biotic farming R^2 is 88% whereas, conventional farming it is 80%. Below Table 3: Performance Metrics of the Random Forest and depicts the results of the models performed. Random forest performs better in both datasets (Table 4 and Fig. 8).

Table 3. Performance metrics of the random forest

Conventional farming (random forest)	Pro biotic farming (random forest)
R2 83%	91%
RMSE 0.2401	1.7475
MAE 0.0833	0.6361

Table 4. Performance metrics of the decision tree

Conventional farming (decision tree)	Pro biotic farming (decision tree)
Accuracy 73%	84%
RMSE 0.6914	0.0484
MAE 0.1937	0.0015

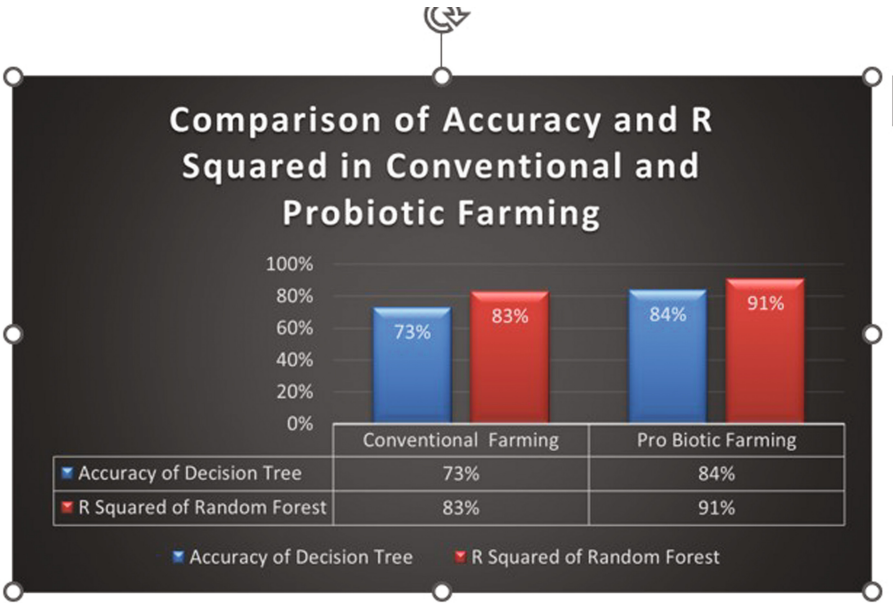


Fig. 8. Comparison of Accuracy and R Squared in conventional farming and probiotic farming

7 Conclusion

In this proposed methodology, a detailed analysis of environmental condition and significant attributes for the growth of the crop is measured and described briefly. Improper monitoring and controlling of crops in traditional farming method fails to result a good yield and fails to maximize the nutritional values in the crop. From the experimental study, it is found that soil electrical conductivity, soil moisture and temperature play's a major role in enhancing NPK in the crop. This review accentuates synergistic characteristics of soil in hybrid farming that could contribute to urban strength, human wellbeing and improved productivity and enhanced nutritional content. With the advancement of precision farming, growth of the crop can be monitored and controlled remotely to enrich the nutritional value. Limitation of the proposed experimental study is the amount of dataset extracted from the dashboard.in future we may extract the dataset for periods of six months to one year. The proposed study carried only for Vallarai (*Centella asiatica*), in future, this methodology will be carried out for all the six crops in Nutri Garden. This proposed study is highly beneficial for researchers and farmers to evaluate the nutritional conduct in the soil with level of soil conductivity in their own soil.

Acknowledgement. The authors of this paper would like to express their sincere gratitude to AAI SP.

Avinashilingam Artificial Intelligence Start-up Programme-CMLI Centre for Machine Learning and Intelligence and for the extended support and guidance to carry out this research work.

References

1. Baweja, P., Kumar, S., Kumar, G.: Fertilizers and pesticides: their impact on soil health and environment. In: Giri, B., Varma, A. (eds.) soil health. SB, vol. 59, pp. 265–285. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44364-1_15
2. Tudi, M., et al.: Agriculture development, pesticide application and its impact on the environment. *Int. J. Environ. Res. Public Health* **18**(3), 1112 (2021)
3. Pahalvi, H.N., Rafiya, L., Rashid, S., Nisar, B., Kamili, A.N.: Chemical fertilizers and their impact on soil health. In: Dar, G.H., Bhat, R.A., Mehmood, M.A., Hakeem, K.R. (eds.) *Microbiota and Biofertilizers, Vol 2: Ecofriendly Tools for Reclamation of Degraded Soil Environs*, pp. 1–20. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-61010-4_1
4. Agricultural Production (2019). Accessed 13 Dec 2022. <https://www.financialexpress.com/economy/india-wastes-up-to-16-of-its-agricultural-produce-fruits-vegetables-squandered-the-most/1661671/>
5. Satterthwaite, D., McGranahan, G., Tacoli, C.: Urbanization and its implications for food and farming. *Phil. Trans. Roy. Soc. B: Biol. Sci.* **365**(1554), 2809–2820 (2010)
6. Sheeja, S., Kabeerathumma, S., Pilla, N.G., Nair, M.M.: Availability and distribution of micronutrients in cassavagrowing soils of Andhra Pradesh. *J. Root Crops* **20**, 75–80 (1994)
7. Pierce, F.J., Nowak, P.: Aspects of precision agriculture. *Adv. Agron.* **67**, 1–85 (1999)
8. Monteiro, A., Santos, S., Gonçalves, P.: Precision agriculture for crop and livestock farming—Brief review. *Animals* **11**(8), 2345 (2021)

9. Sharma, A., Jain, A., Gupta, P., Chowdary, V.: Machine learning applications for precision agriculture: a comprehensive review. *IEEE Access* **9**, 4843–4873 (2020)
10. de Souza Vandenbergh, L.P., et al.: Potential applications of plant probiotic microorganisms in agriculture and forestry. *AIMS Microbiol.* **3**(3), 629 (2017)
11. Reddy, S.P., Sneha, B., Vinothini, M., Kritthika, R.V., Deepika, Y.: IoT based smart precision agriculture in rural areas. *Eur J Mol Clin Med* **7**(4), 1443–1451 (2020)
12. Andrew, R.C., Malekian, R., Bogatinoska, D. C.: IoT solutions for precision agriculture. In: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 0345–0349 (2020)
13. Heiniger, R.W., McBride, R.G., Clay, D.: E: Using soil electrical conductivity to improve nutrient management. *Agron. J.* **95**(3), 508–519 (2003)
14. Cen, Y., Guo, L., Liu, M., Gu, X., Li, C., Jiang, G.: Using organic fertilizers to increase crop yield, economic growth, and soil quality in a temperate farmland. *PeerJ* **8**, e9668 (2020)
15. Hussein, H.S., Shaarawy, H.H., Hussien, N.H., Hawash, S.I.: Preparation of nano-fertilizer blend from banana peels. *Bull. Natl. Res. Centre* **43**(1), 1–9 (2019). <https://doi.org/10.1186/s42269-019-0058-1>
16. Song, Y.Y., Ying, L.U.: Decision tree methods: applications for classification and prediction. *Shanghai Arch. Psychiat.* **27**(2), 130 (2015)
17. Chai, T., Draxler, R.: R: Root mean square error (RMSE) or mean absolute error,(MAE). *Geosci. Model Dev. Disc.* **7**(1), 1525–1534 (2014)
18. Biau, G., Scornet, E.: A random forest guided tour. *TEST* **25**(2), 197–227 (2016). <https://doi.org/10.1007/s11749-016-0481-7>
19. ElNour, M.E., Elfadil, A.G., Manal, F.A., Saeed, B.: A: Effects of banana compost on growth, development and productivity of sorghum bicolor cultivar (Tabat). *J. Adv. Biol.* **8**(2), 1–7 (2015)
20. Aboul-Enein, A.M., Salama, Z.A., Gaafar, A.A., Aly, H.F., Abou-Elella, F., Ahmed, H.: A: Identification of phenolic compounds from banana peel (*Musa paradaisica* L.) as antioxidant and antimicrobial agents. *J. Chem. Pharmaceut. Res.* **8**(4), 46–55 (2016)



Evaluation and Comparison of Energy Consumption Prediction Models Case Study: Smart Home

Elhabyb Khaoula^(✉), Baina Amine, and Bellafkih Mostafa

RAISS Lab, National Institute of Posts and Telecommunications - INPT Rabat,
Rabat, Morocco

{elhabyb.khaoula,baina,bellafkih}@inpt.ac.ma
<http://www.inpt.ac.ma/>

Abstract. Predicting a structure's energy usage is an essential part of achieving energy efficiency objectives. Engineering, AI-based, and hybrid approaches can all be used to predict how much energy a building will require; however, we choose the AI-based method because it uses historical data to make predictions about future energy usage rather than thermodynamic equations the other approaches rely on. As a result, the objective of this study is to put several prediction models for energy usage into practice and assess them, the recommended algorithms are linear regression, random forest, and artificial neural network. Our study's data set was gathered from a house that served as a case study, and we compared each approach's efficacy using RMSE, R squared, MAE, and MAPE measurements.

Keywords: Smart house · Machine learning · Home energy use prediction · Supervised learning · Intelligent buildings

1 Introduction

The notion of interconnected infrastructures and machines, which may be seen as having its oldest roots in the year 1923, is where the topic of building intelligence and the learning capacity of buildings begins. The architect Le Corbusier described a house as “a machine for living in” at that time [1]. In order to address the rising ability of artificial systems to act autonomously, computer science eventually led to the introduction of the phrase “intelligence.” At the same time “smart buildings” were first mentioned in the scientific literature in the 1990s (Derek et al. [2]). The necessity to illustrate the relationship between humans and robots was realized at the start of the 2000s. For instance, building intelligence may refer to knowledge “imprinted into an artificial object (such as a structure) by human intelligence,” according to Himanen (2003) [3].

Generally speaking, the term “smart building,” sometimes known as “intelligent building,” refers to the integration of digital technology into buildings intended for residential or commercial use [4]. The major goal is to reduce a building's energy use while improving the comfort of its occupants. So, thanks to its

attributes, the intelligent building responds to today's environmental concerns. Currently, smart buildings work according to the principle of smart grids implemented in electricity distribution networks based on technological advances, these intelligent grids are able to identify peaks in energy consumption and adjust the production and distribution of electricity accordingly, in order to avoid energy waste [5]. So, it is clear now that the biggest benefit of intelligent buildings is their energy consumption, which is controlled at all times to be as close as possible to the requirements of their occupants.

There is an indicator called SIS (System Intelligence Score) [6] that may be used to evaluate the building's level of intelligence, this score has five levels but the most important one is the predictive level because it relates to the capacity to forecast and predict the building's response to future changes in the environment, circumstances of usage, and prospective expectations. So, any smart infrastructure must have a smart strategy that forecasts energy consumption as a power-saving approach since the prediction is the highest degree of the intelligence score system [7]. This appears to be the case since it offers the benefit of enhancing economic returns and acts as a sustainable approach to energy management to decrease energy waste. Researchers have developed a range of modeling strategies to predict building energy consumption since the 1990s. Engineering approaches, AI-based methods, and hybrid methods are further classifications for these technologies [8]. In our case, we are interested in AI-based technologies because it predicts and estimates energy consumption without knowing the underlying connections between the building and its many components, thanks to that the prediction of a structure is generally based on the local environment and the building's attributes. The main goal of this study is to compare various energy consumption forecasting models for smart buildings before implementing them. To do this, we'll use machine learning techniques like linear regression, random forest, and neural network models to analyze historical data sets made up of daily electricity consumption in a home located at Stambruges, a Belgian city 24km away from the City of Mons. The paper is structured as follows: Sect. 1 introduces the idea of energy forecasting for intelligent buildings and homes. Section 2 provides an overview of current advancements in AI-based building energy use prediction. An implementation of the selected models is presented in Sect. 3. Results and analysis are presented in Sect. 4. Conclusions are discussed in Sect. 5.

2 Literature Review

Scientific research into artificial intelligence led to the invention of the powerful machine learning approach, which has enormous development potential. With the right model and methods, machine learning might "learn" the nonlinear connection between the independent variables and the target variables based on historical data. As smart houses are an instance of a smart building, we will present some background information on energy prediction on smart buildings and then on smart houses. Generally speaking, in the field of smart buildings exists different types of buildings, like example Gassar et al. [9] worked on a Residential

building, and in their evaluation, many machine learning methods for predicting electricity and gas demand were implemented. The input features investigated in the study were economic, socio-demographic, and building factors, and it took into consideration the multi-layer neural network (MNN), recurrent neural network (RNN), random forest, and gradient boosting (GB) approach, according to the findings, household income, the number of households, and building characteristics are what influence gas and electricity use the most. For smart homes we take as an example Nicoleta et al. [13] their major goal was to predict the energy consumption of the next 24 h, they used Two basic predictors Decision tree and Artificial neural networks on historical data of homes in France, at the end they suggest two processing to improve the performance which is the aggregation and segmentation of the data. We provide further details regarding each idea of work in Table 1 along with a description of the prediction models that were used, as well as the type of structure, the predicted time scale, and the assessment methodologies.

3 Methodology

By the use of historical data, the AI-based method forecasts the trends in the demand for energy. A procedure with four primary parts, data collecting, data pre-processing, model training process, and model testing, it takes into account the effects of important variables such as building attributes and ambient circumstances. This research aims to predict energy consumption using a time-series data set gathered from a house for four months. This data set is composed of 28 variables such as the temperature and humidity in every room inside and outside the house, the light energy consumption, the appliance energy consumption, wind speed, visibility, date, pressure, and temperature from the chievere station that was acquired from the Reliable Prognosis which is used to compare between the outside properties of the house. Our first test (currently publishing) was on a commercial building with data gathered from 2017 to 2019, we implement four machine learning algorithms to predict electricity consumption, and as we already mentioned houses are an instance of buildings, that's the reason why we tested the forecasted energy consumption on an individual structure to better understand what features will be usable for a good prediction of energy consumption for buildings and then for grids. The prediction techniques that would be used are Artificial Neural Network, Random Forest, and Simple Linear Regression. The raw data will be first examined and pre-processed, and the validation measures will then be used to assess each model. This would reduce the complexity of the process of the model training and handle any missing data. The Four steps of our work are:

1. Exploratory data analysis
2. Features engineering
3. Prediction algorithms
4. Model evaluation.

Table 1. Machine learning background for smart constructions

	Type	Used algorithm	Prediction time scale	Evaluation measurement	Concept
Smart buildings	Residential building	MNN RNN GBM RF	2 Years	MAE MAPE MSE	Gasser et al. [9] In order to forecast gas and electricity usage, they have implemented the multi-layer neural network, neural networks as well as random forest and gradient boosting approaches
	Residential building	SVR MLR ANN	15 months	RMSE MAE MAPE	Massana et al. [10] predicted the short-term load for non-residential buildings using SVR, MLR, and ANNs
	Commercial building	ANN	10 months	MAPE	Chae et al. [11] Integrated an ANN model with a Bayesian regularization approach to forecasting sub-hourly power demand in commercial buildings
Smart homes	Simple home	DT ANN	NL	RMSE MAPE	RV.jones et al. [12] Used decision trees and neural networks to examine the appliances influencing the use of electricity in houses
	Simple home	RF BN	7 months	MAPE RMSE	Nicoleta et al. [13] In order to predict the energy consumption in houses for the next 24 h, they used random forests and an Artificial neural network to analyze the historical data set and create a prediction model
	Low energy house	GMB RF SVR	6 months	RMSE R-square MAE MAPE	In order to predict the energy consumption of a low-energy house Candanedo et al. [14] used diverse statistical models and the GBM gives the best precision at 97%

3.1 Exploratory Data Analysis

In this part, we verify if there is missing data by applying the missingness matrix, then detecting the outliers by arranging the data-set point and calculating the median and after calculating the upper and lower quartile and as the last operation calculate the inner fences for the data-set, as a result of the operation we have excluded around 15% of the data from the formulas above, which is fine.

3.2 Features Engineering

After the first step, we remarked that Week Status and Days of Week are two columns with the potential for modification. There are no null values in the data set, and the date column is not necessary because the NSM column has already been interpreted, based on this we transform the Week-Status and the Days of Week, then we redefine the appliance column.

3.3 Prediction Algorithms

To anticipate energy use, this study used supervised machine learning techniques. Data were partitioned into two groups, a training group made up of 70% of the data and a testing group made up of 30%, before being used to develop and train the model. The algorithms utilized are shown above:

Random Forest: To implement the random forest approach, a decision tree is trained using a new bootstrap sample from the training set, at each node of the decision tree, features are chosen at random. Then, the Gini impurity (a measurement of the likelihood of incorrectly classifying a new instance of a random variable) is applied only to that set of features [15]. The best feature is then chosen, and the process is repeated until the tree is complete. For regression, the random forest prediction is the unweighted average over the collection [16]:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + e_i \quad (1)$$

Artificial Neural Network: ANNs are frequently used to predict electricity use as well as the cooling and heating loads in buildings. An artificial neural network with several layers is referred to as a “deep neural network” and is one sort of deep learning model [17] and it is characterized by the next formula:

$$f(b + \sum_{i=1}^n (x_i u w_i)) \quad (2)$$

with b bias, x input of neuron, n : the number input from the incoming layer, and i a counter from 1 to n .

Simple Linear Regression: is a case model with a single independent variable. Simple Linear regression defines the dependence of the variable, and Simple regression distinguishes the influence of independent variables from the interaction of dependent variables [18].

$$y = \beta_0 + \beta_1 + \epsilon \quad (3)$$

3.4 Model Evaluation (Test)

The performance and accuracy of each machine learning algorithm’s predictive model were evaluated after the construction and production of prediction demand data. Root Mean Square Error (RMSE), R squared, Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) metrics are used to compare the performance of each approach. Table 2 presents each metric. Where Y_i is the actual measurement (energy consumption), Y'_i is the predicted value and n is the number of measurements.

Table 2. Performance metrics

Algorithms	Description
R Squared [19]	$R^2 = 1 - \frac{\sum_{i=1}^{n-1} (Y_i - Y'_i)^2}{\sum_{i=1}^{n-1} (Y_i - Y''_i)^2}$
Root Mean Square Error [19]	$RMSE = \sqrt{\frac{\sum_{i=1}^{n-1} (Y_i - Y'_i)^2}{n}}$
Mean Absolute Error [20]	$MAE = \frac{\sum_{i=1}^{n-1} (Y_i - Y'_i)^2}{n}$
Mean Absolute Percentage Error [20]	$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - Y'_i)^2}{n}$

4 Result and Discussions

This part examined the experiment's findings in light of the anticipated progression. A performance comparison of the forecasts from Random Forest (RF), Simple Linear Regression (SLR), and artificial neural network (ANN) was provided after the energy consumption forecast data were analyzed.

The effectiveness of the suggested prediction approaches was therefore tested after model training and testing. By analyzing the efficiency of each approach, the result of the training and testing could be seen in Table 3. We began with the testing data-set, Random Forest gives the best correlation coefficient (0.95), and also the less error tau: RMSE = 26%, MAE = 30%, and MAPE = 29.34% which shows that the model generates a minimum number of errors. In the second place, LR offers 63% of correlation, and a medium error tau of 39%, 52%, and 62% respectively from RMS, MAE, and MAPE. Lastly, Artificial neural networks give a negative correlation coefficient of 0,17% and the biggest error numbers. So, we can conclude that Random Forest is the best prediction model for the testing data set. In term of training data, there is no big difference Random Forest offer a precision of 88%, Less error number again with 23%, 10%, and 10% respectively RMSE, MAE and MAPE, secondly comes LR and finally ANN. Table 3 shows that RF, has excellent precision, and it gets the closest to the actual testing values, followed by LR and ANNA. As shown in Fig. 1, the comparison of the

Table 3. Performance evaluation predictions using trained models

Measurements	LR	RF	ANN
R Squared Test	0.63	0.88	-0.07
RMSE Test	39.50	26.18	107.27
MAE Test	53.83	30.13	48.61
MAPE Test (%)	62.02	29.34	36.74
R Squared Train	0.45	0.95	-0.06
RMSE Train	52.57	23.59	104.48
MAE Train	32.07	10.89	46.45
MAPE Train (%)	26.58	10.82	36.42

average consumption between the actual and projected values was tabulated and shown as a line graph.

As a conclusion of the results from the table and figure, we conclude that the best performances are given by Random Forest and the reason why is that it's a decision tree algorithm that selects the most influential features to make a good prediction. We already tried the Longest short-term memory, Multiple Linear Regression, and Random Forest as a first try, and from the two tests, we conclude that each algorithm has its properties but those that give good precision are time series algorithm and decision-tree-based, also the type of the data plays a major role in algorithms precision that's why we tried a time series data.

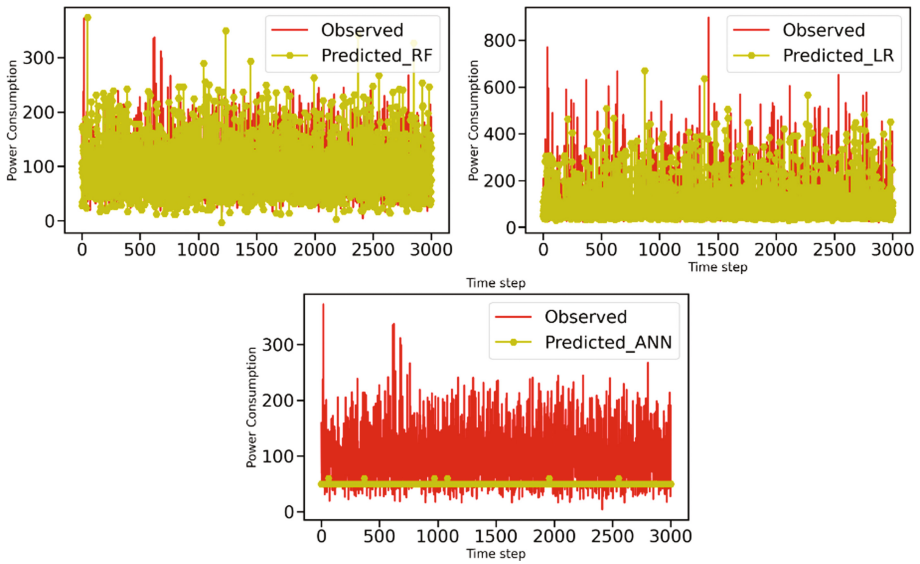


Fig. 1. Comparison between the actual and forecasted average consumption

5 Conclusion

The objective of this study is to examine several energy forecast models for a house. The four-month-old energy demand data was analyzed and made ready for the prediction model training and testing process. A statistical study of the data was done to verify the dataset's normality, and the results showed that the distributional characteristics of all the collected data is normal. Three supervised machine learning prediction approaches were used to forecast the energy consumption in the house, which are Linear Regression, Random Forest, and artificial neural network. Effective, comparisons were made between these techniques' final structures and prediction performance. The outcomes of the model training and testing showed that each tactic performed differently. The

most positive outcomes were from the Random Forest algorithm, which had accuracy values of 95.3% and 88.8% for training and testing data, respectively. With a few suggestions for further research, this work might be enhanced. Since a hybrid or ensemble technique performs better than a single predictor, it can be advised. Additionally, several different types of data sets must be tested in order to find the best approach from various perspectives.

Acknowledgement. The authors would like to thank the National Center for Scientific and Technical Research (CNRST) for supporting and funding this research.

References

1. Le Corbusier, C.-E.J.: Vers une architecture (In french: Towards an architecture), collection de "l'esprit nouveau", les 'EDITIONS g. Paris, France: CR'ES ET C (1923)
2. Clements-Croome, T., Derek, J.: What do we mean by intelligent buildings? *Autom. Constr.* **6**(5), 395–400 (1997)
3. Himanen, M.: The intelligence of intelligent buildings: the feasibility of the intelligent building concept in office buildings. VTT Technical Research Centre of Finland (2003)
4. Xie, X., Lu, Q., Herrera, M., Yu, Q., Parlikad, A.K., Schooling, J.M.: Does historical data still count? Exploring the applicability of smart building applications in the post-pandemic period. *Sustain. Cities Soc.* **69**, 102804 (2021)
5. Wang, Z., Liu, J., Zhang, Y., Yuan, H., Zhang, R., Srinivasan, R.S.: Practical issues in implementing machine-learning models for building energy efficiency: moving beyond obstacles. *Renew. Sustain. Energy Rev.* **143**, 110929 (2021)
6. Wong, J., Li, H., Lai, J.: Evaluating the system intelligence of the intelligent building systems: Part 1: development of key intelligent indicators and conceptual analytical framework. *Autom. Constr.* **17**(3), 284–302 (2008)
7. Xu, D., et al.: A classified identification deep-belief network for predicting electric-power load. In: 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), pp. 1–6. IEEE (2018)
8. Fouquier, A., Robert, S., Suard, F., St'ephan, L., Jay, A.: State of the art in building modelling and energy performances prediction: aD review. *Renew. Sustain. Energy Rev.* **23**, 272–288 (2013)
9. Gassar, A.A.A., Yun, G.Y., Kim, S.: Data-driven approach to prediction of residential energy consumption at urban scales in London. *Energy* **187**, 115973 (2019)
10. Massana, J., Pous, C., Burgas, L., Melendez, J., Colomer, J.: Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes. *Energy Build.* **130**, 519–531 (2016)
11. Chae, Y.T., Horesh, R., Hwang, Y., Lee, Y.M.: Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy Build.* **111**, 184–194 (2016)
12. Jones, R.V., Fuertes, A., Lomas, K.J.: The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. *Renew. Sustain. Energy Rev.* **43**, 901–917 (2015)
13. Arghira, N., Hawarah, L., Ploix, S., Jacomino, M.: Prediction of appliances energy use in smart homes. *Energy* **48**(1), 128–134 (2012)

14. Candanedo, L.M., Feldheim, V., Deramaix, D.: Data driven prediction models of energy use of appliances in a low-energy house. *Energy Build.* **140**, 81–97 (2017)
15. Wang, Z., Wang, Y., Zeng, R., Srinivasan, R.S., Ahrentzen, S.: Random Forest based hourly building energy prediction. *Energy Build.* **171**, 11–25 (2018)
16. Segal, M.R.: Machine learning benchmarks and random forest regression (2004)
17. Ahmad, M.W., Mourshed, M., Rezgui, Y.: Trees vs Neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy Build.* **147**, 77–89 (2017)
18. Maulud, D., Abdulazeez, A.M.: A review on linear regression comprehensive in machine learning. *J. Appl. Sci. Technol. Trends* **1**(4), 140–147 (2020)
19. Kenney, J.F.: Mathematics of statistics. D. Van Nostrand (1939)
20. Everitt, B.S., Skrondal, A.: The Cambridge dictionary of statistics (2010)



Leaf Disease Detection in Blueberry Using Efficient Semi-supervised Learning Approach

Vinh Dinh Nguyen¹(✉), Ngoc Phuong Ngo², and Narayan C. Debnath¹

¹ School of Computing and Information Technology, Eastern International University,
Binh Duong, Vietnam

{vinh.nguyen,narayan.debnath}@eiu.edu.vn

² Biochemistry and Plant Physiology Faculty, College of Agriculture,
Can Tho University, Can Tho City, Vietnam

npngoc@ctu.edu.vn

Abstract. Blueberry leaf disease detection is really important to help farmers to early detect leaf disease and find a suitable method to cure the disease. Therefore, this research introduces an approach to detect and classify blueberry leaf disease by using an unsupervised method (auto-encoder), and a supervised method (support vector machine). The accuracy of our proposed method was evaluated by conducting the experiments on blueberry dataset captured at Can Tho City, Vietnam. The existing augmentation techniques was also applied to increase the data size of training and testing. For the first experiment on normal capturing conditions, the F1 scores of the proposed method and SVM are 89.28% and 81.48%, respectively. For the second experiment with noisy conditions, the F1 scores of the proposed method and SVM are 81.5% and 66.7%, respectively.

Keywords: Blueberry leaf disease · Deep learning · Un-supervised learning · Unsupervised learning

1 Introduction

Nowadays, leaf disease is one of the typical issues that affect farmers. Therefore, several searchers did a study to find a way to help farmers by developing a method that automatically detects and classifies leaf disease, automatically. The researcher often uses a supervised learning approach to detect and classify blueberry leaf diseases such as SVM, Random Forest, Neural Networks, and Deep learning as discussed [1]. Most of the existing blueberry leaf disease detection is designed by allowing the raw input pixel to the proposed method for training and testing [1]. However, in this research, the benefit of unsupervised learning was used to extract stable groups of features before inputting them into the SVM for train and testing.

To increase the accuracy of the existing support vector machine (SVM)-based method for object detection and classification, the authors study the advantages of the autoencoder [2]. First, the first proposed auto-encoder is used to extract and reduce the features dimension of the input image size from $160 \times 160 \times 3$ to 100 features. Second, the second auto-encoder is used to extract and reduce the feature dimension of the input image

resolution from $160 \times 160 \times 3$ to 200 features. Third, the third auto-encoder is used to extract and reduce the feature dimension of the input image resolution from $160 \times 160 \times 3$ to 300 features. Fourth, the results of 100, 200, and 300 features are concatenated to establish 600 features for input to the SVM to detect and classify whether the input blueberry leaf has a disease or not.

The remaining paper was organized as follows: Sect. 2 summarizes the existing research in Blueberry leaf disease detection. Section 3 introduces and discusses the main components of the proposed method. Section 4 shows and discusses the accuracy of the proposed method under various testing conditions. Section 5 mentions the limitation and future work.

2 Related Works

To detect blueberry leaf disease, various machine learning and deep learning approaches have been investigated. Barbedo et al. developed a method to detect and classify various blueberry disease, such as citrus caker, bacterial blight, and black mouth, by using the convolutional neural network [3, 4]. Fuentes et al. introduce a deep learning approach to detect various diseases of tomato plants by using SSD and Faster RCNN [5]. Lee et al. developed a new CNN along with a pre-train VGG model to accurately detect and classify leaf disease by using an open dataset [6]. Another research was conducted by Too et al. [7] on black rot and early blight by using DenseNet and RestNet. Recently, in the AI-challenge plant disease contest, Zhong et al. use DenseNet-121 to recognize apple leaf disease [8]. More recently, Shrivastava et al. developed an improved version of CNN and SVM to detect and classify Rice plants [9, 10].

3 The Proposed Method

To increase the accuracy of the existing support vector machine (SVM)-based method for object detection and classification, the authors study the advantages of the auto encoder [2]. First, the first proposed auto-encoder is used to extract and reduce the features dimension of the input image size from $160 \times 160 \times 3$ to 100 features. Second, the second auto-encoder is used to extract and reduce the feature dimension of the input image resolution from $160 \times 160 \times 3$ to 200 features. Third, the third auto-encoder is used to extract and reduce the feature dimension of the input image resolution from $160 \times 160 \times 3$ to 300 features. Fourth, the results of 100, 200, and 300 features are concatenated to establish 600 features for input to the SVM to detect and classify whether the input blueberry leaf has a disease or not (as shown in Fig. 1).

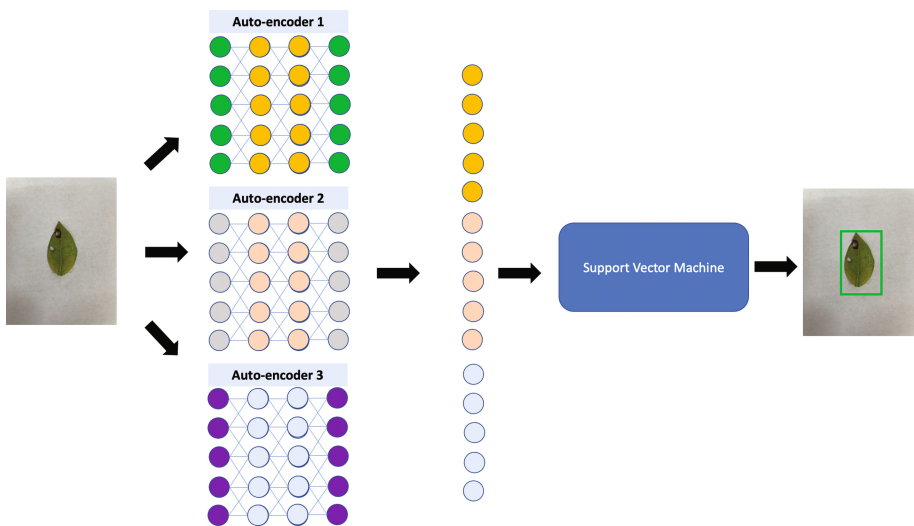


Fig. 1. The auto-encoder is used to generate three robust groups of features for training support vector machine [2]

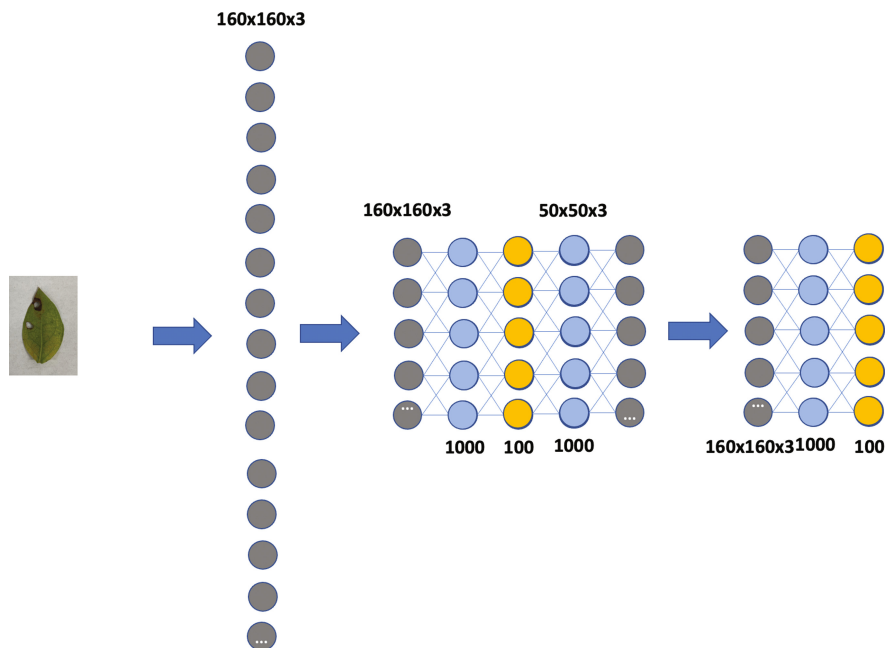


Fig. 2. The first proposed auto-encoder to learn and reduce input features ($160 \times 160 \times 3$) to 100 features.

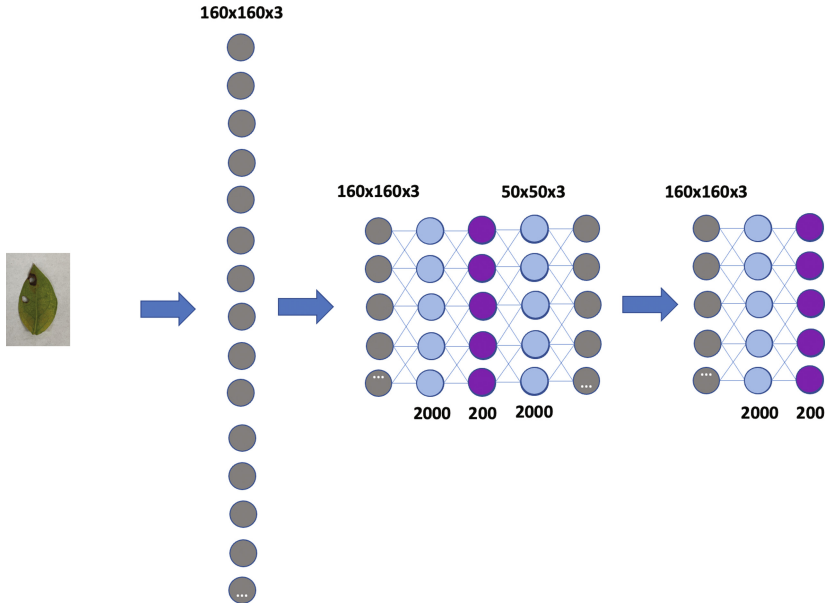


Fig. 3. The second proposed auto-encoder to learn and reduce input features ($160 \times 160 \times 3$) to 200 features.

The first proposed au-encoder was designed by considering the input image of $160 \times 160 \times 3$ pixels (as shown in Fig. 2). The first hidden layer was designed with 1000 units. The second hidden layer was designed with 100 units. The third hidden layer was designed with 1000 units and the output layer was designed with the same units as the input layer.

The second proposed au-encoder was designed by considering the input image of $160 \times 160 \times 3$ pixels (as shown in Fig. 3). The first hidden layer was designed with 2000 units. The second hidden layer was designed with 200 units. The third hidden layer was designed with 2000 units and the output layer was designed with the same units as the input layer.

The third proposed au-encoder was designed by considering the input image of $160 \times 160 \times 3$ pixels (as shown in Fig. 4). The first hidden layer was designed with 3000 units. The second hidden layer was designed with 300 units. The third hidden layer was designed with 3000 units and the output layer was designed with the same units as the input layer.

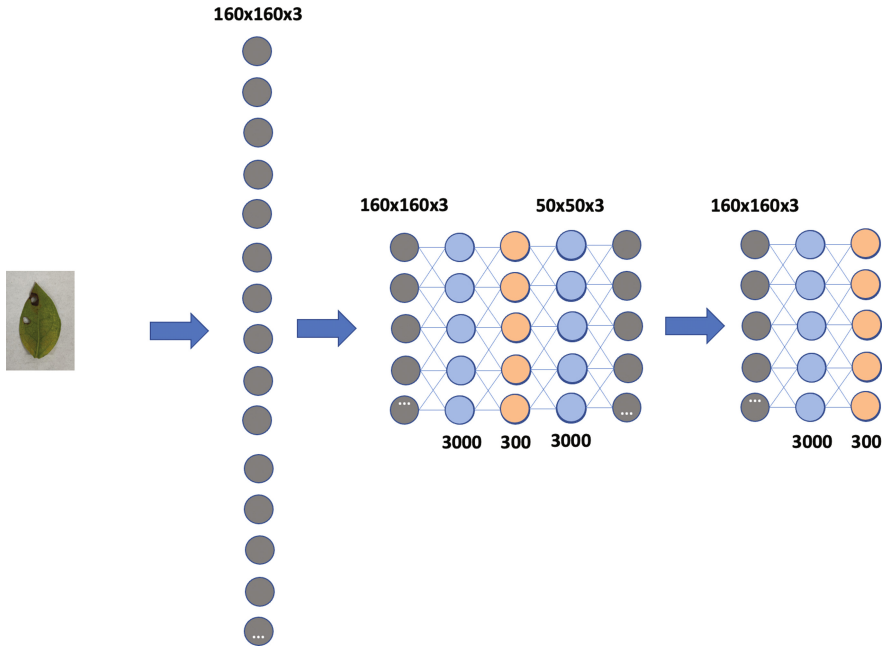


Fig. 4. The third proposed auto-encoder to learn and reduce input features ($160 \times 160 \times 3$) to 300 features.

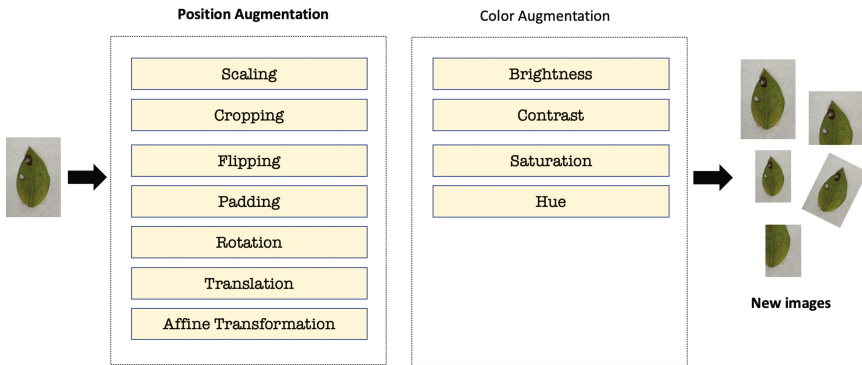


Fig. 5. Flow chart of generating images for training and testing using the augmentation techniques [11]

4 Experimental Results

4.1 Blueberry Dataset for Training and Testing

A cell phone Samsung Galaxy A33 5G was setup to capture and label blueberry images for training and testing. The dataset was built at Can Tho University, Vietnam. The total number of images captured and labeled is 200 images. However, the captured

images were still not enough to train and evaluate the accuracy of the proposed method. Various common augmentation techniques [11] were applied to generate 1000 images for training, 200 images for testing, and 200 images for validation as shown in Fig. 5.

4.2 Evaluation Metric and System Configuration for Training and Testing

The proposed system was trained by using free GPU Google CoLab provided by Google. The accuracy of the proposed method with the original support vector machine (SVM) [1] were evaluated by using three popular metrics: precision., recall, and F1 score.

4.3 Results

This research is aim to detect and classify whether the blueberry has a disease or not by using our captured dataset at Can Tho University, Vietnam. The first experiment was conducted by evaluating the performance of our system with the SVM under normal testing conditions as shown in Fig. 6. Blueberry leaf disease was successfully detected by our proposed method which is much better than the results from the SVM in this test. The blueberry leaf disease detection was also evaluated under more challenging conditions by adding salt and pepper noise to the input image described in Fig. 7. The results from the proposed method are still stable in comparison to the results from the SVM in this experiment.

To future verify the effectiveness of the proposed method and SVM under large datasets, 200 images under normal environments and noisy environments was used to evaluate the performance of compared methods. Figure 8 provides the precision, recall,

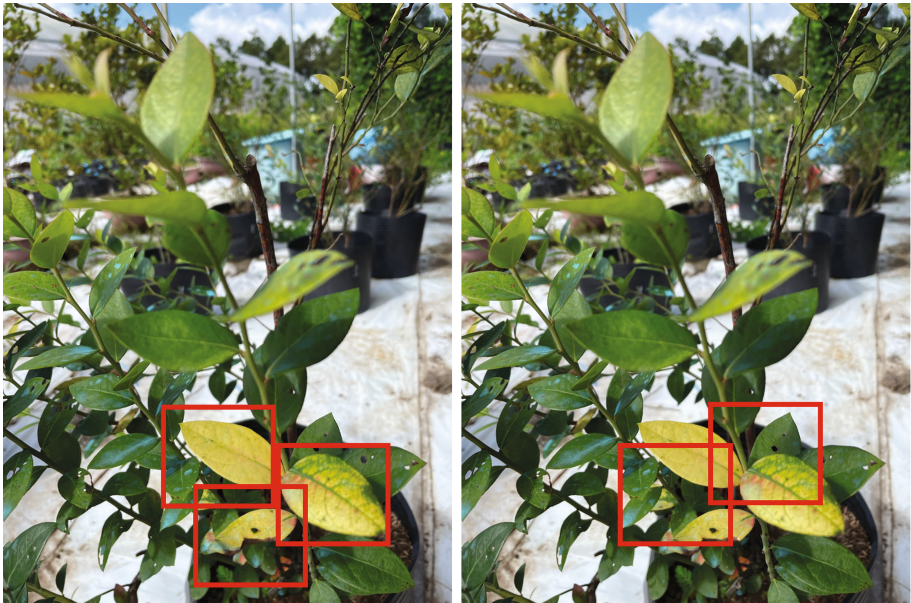


Fig. 6. Blueberry detection results by using our dataset under normal condition. Left image is the results of the proposed method. Right image is the results of the SVM.

and F1 of the proposed method and SVM under normal testing conditions. In this test, the F1 scores of the proposed method and SVM are 89.28% and 81.48%, respectively. Figure 9 provides the precision, recall, and F1 of the proposed method and SVM under difficult testing conditions by adding noise to the input image. In this test, the F1 scores of the proposed method and SVM are 81.5% and 66.7%, respectively.

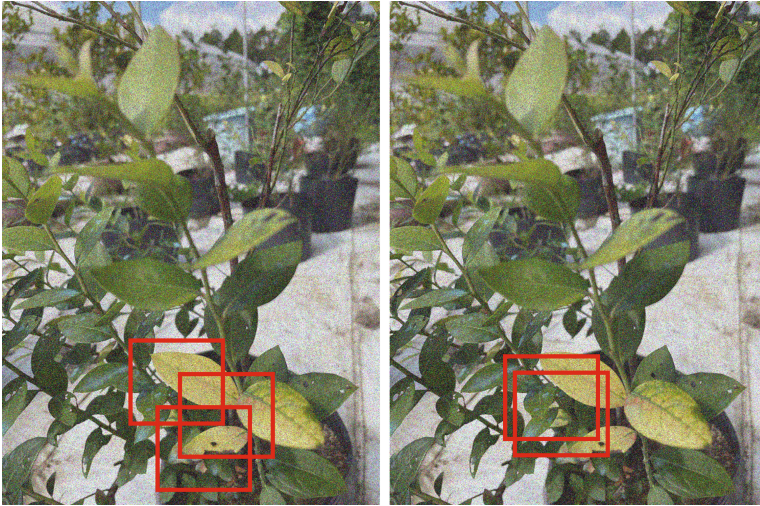


Fig. 7. Blueberry detection results by using our dataset under noisy conditions. Left image is the results of the proposed method. Right image is the results of the SVM.

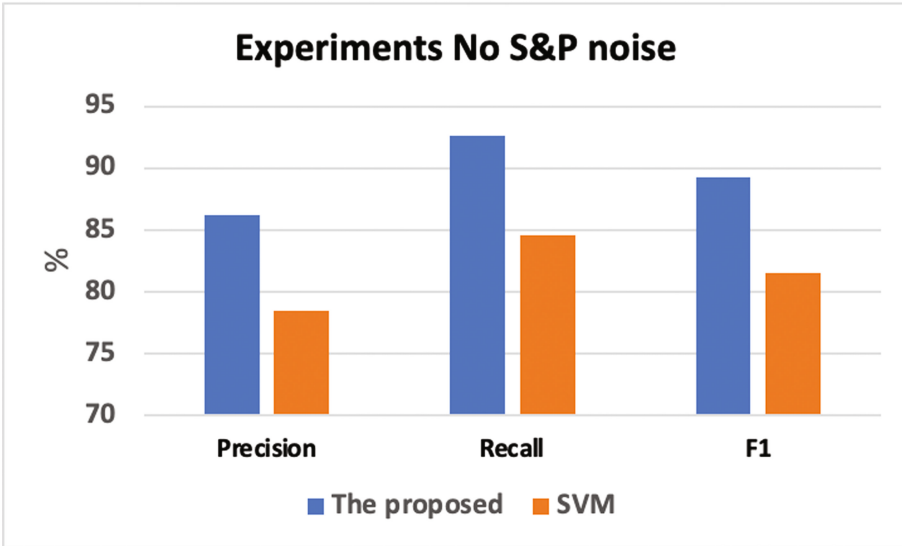


Fig. 8. Precision and Recall of the proposed method and support vector machine under the testing condition without adding Salt and Pep noise

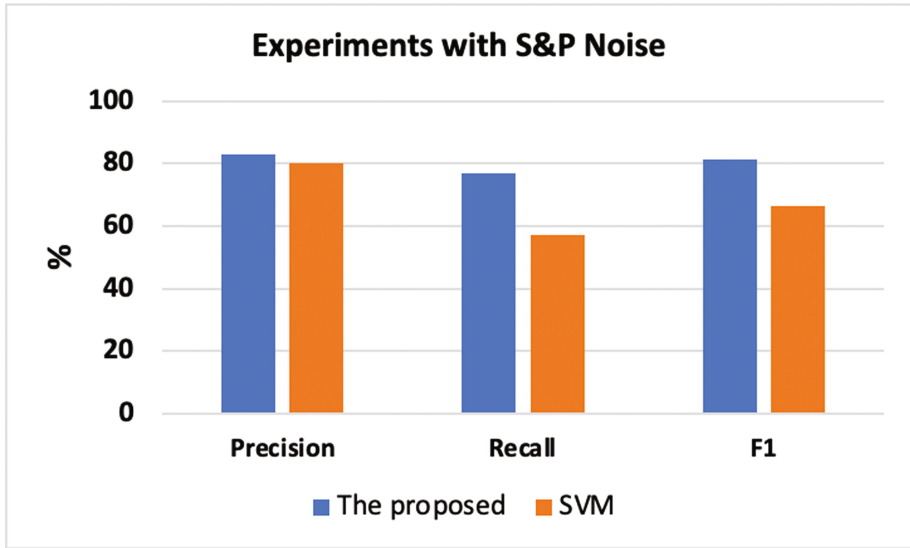


Fig. 9. Precision and Recall of the proposed method and support vector machine under the testing condition by adding salt and pepper noise

5 Conclusion

Blueberry leaf disease detection is really important to help farmers to early detect and find a way to reduce the cost. The proposed method, by combining the unsupervised method (auto-encoder) and supervised method (SVM), can detect and classify blueberry leaf disease under various testing environments. However, the processing time of our system is still slow due to the proposed auto-encoder for learning and extracting three groups of features. Therefore, in the future, The effect of various auto-encoder variants are investigated in order to improve the performance of our system.

Acknowledgments. The authors would like to express gratitude to Eastern International University (EIU), Binh Duong, Vietnam, for funding this research.

References

1. Cecilia, S., Carlos, M., Carlos, R., Thais, F.: Diseases detection in blueberry leaves using computer vision and machine learning techniques. *Int. J. Mach. Learn. Comput.* **9**(5), 656–661 (2019)
2. Zhou, J., Ju, L., Zhang, X.: A hybrid learning model based on auto-encoders. In: 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 522–528 (2017)
3. Barbedo, J.G.A.: Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* **153**, 46–53 (2018)
4. Militante, S.V., Gerardo, B.D., Dionisio, N.V.: Plant leaf detection and disease recognition using deep learning. In: Proceedings of the 2019 IEEE Eurasia Conference on IOT,

Communication and Engineering (ECICE), Yunlin, Taiwan, 3–6 October 2019, pp. 579–582 (2019)

5. Fuentes, A., Yoon, S., Kim, S.C., Park, D.S.: A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* **17**, 2022 (2017)
6. Lee, S.H., Goëau, H., Bonnet, P., Joly, A.: New perspectives on plant disease characterization based on deep learning. *Comput. Electron. Agric* **170**, 105220 (2020)
7. Too, E.C., Yujian, L., Njuki, S., Yingchun, L.: A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric* **161**, 272–279 (2019)
8. Zhong, Y., Zhao, M.: Research on deep learning in apple leaf disease recognition. *Comput. Electron. Agric* **168**, 105146 (2020)
9. Shrivastava, V.K., Pradhan, M.K., Minz, S., Thakur, M.P.: Rice plant disease classification using transfer learning of deep convolution neural network. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **3**, 631–635 (2019)
10. Shrivastava, V.K., Pradhan, M.K.: Rice plant disease classification using color features: a machine learning paradigm. *J. Plant Pathol* **103**, 17–26 (2021)
11. Rrmoku, B., Qehaja, B.: Data augmentation techniques for expanding the dataset in the task of image processing. In: 29th International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 1–3 (2022)



Student Performance Prediction in Learning Management System Using Small Dataset

Zakaria Soufiane Hafdi and Said El Kafhali^(✉) 

Faculty of Sciences and Techniques, Computer, Networks, Modeling, and Mobility
Laboratory (IR2M), Hassan First University of Settat, 26000 Settat, Morocco
{z.hafdi,said.elkafhali}@uhp.ac.ma

Abstract. Predicting students' performance has become a major need in most educational institutions. This is necessary to support at-risk students, ensure their retention, provide top-notch learning opportunities, and improve the university's efficiency and competitiveness. Even so, this may be difficult to collect records for medium-sized institutions., particularly those that concentrate on graduate and postgraduate programs and have a limited number of applicant records available for examination. Therefore, the prime objective of this research is to demonstrate the viability of constructing and training a predictive model with a credible accuracy rate using a modest dataset size. This study also investigates the possibilities of employing, visualization and clustering techniques to identify the key factors in the set of data used to build classification models. The most accurate model was determined by evaluating the best indications through various machine-learning methods.

Keywords: Machine learning · Learning analytics · E-Learning · Learning management system

1 Introduction

Numerous efforts have been made to forecast student performance to achieve a variety of goals, such as identifying at-risk students, ensuring student retention, allocating resources programs, and many others. Early predicting factors may provide at-risk students with the opportunity to receive a high-quality education. Machine learning algorithms is used for that purpose in studies is part of the field of Educational Data Mining (EDM) [1].

Additionally, the adoption of new and improved technologies in educational systems allows some students to pursue their educational goals. To reach this goal, it is necessary to revisit previously disregarded concepts and technologies. For example, Learning Management Systems (LMS) [2] have fallen out of favor in recent years because some institutions regard them as just repositories. But after Corona Decease, this vision has taken a drastic turn, with LMS serving students to retain engagement with their institutions [3].

It is our responsibility as researchers to provide new ways to explore educational data to determine the efficacy of learning systems, analyze the academic performance of students, and create an early warning system. However, it remains a difficult task due to the impact of various elements that affect student performance [4], such as family factors, psychological profile, past schooling, previous academic success, and contact between students with peers and teachers.

This study aims to forecast student performance in order to involve different learners in studies and projects that may increase institutions' reputations and rankings both locally and internationally.

Additionally, in the majority of studies that sought to categorize or forecast, researchers needed to put in a lot of work merely to isolate the key variables that would be most helpful in creating plausible and accurate prediction models. However, feature ranking methods, looking at the selected features, and extracting patterns are the most applied by machine learning algorithms in the training phase [5].

Instead, and up until lately, there have not been any research attempts to look into how visualization or clustering approaches would be able to discover these factors for our data case, especially in the EDM field, in this work [6]), the authors studied the impact of focusing a particular number of courses to get best results.

By answering the following research questions, this study hopes to close the aforementioned gaps.

- What machine learning classification model has the highest accuracy and the smallest dataset size for classifying a student's dissertation project grade?
- What are the main factors that predict students' performance (as measured by course grades and dissertation grades)?

To address these questions, the main contribution of this study is to demonstrate the feasibility of building and training a prediction model with an acceptable accuracy rate using a small dataset size.

The remainder of this paper is structured as follows. The related work will be discussed in Sect. 2. The dataset will be described in Sect. 3. The section will show the results of the analysis and discussion Sect. 4. At last, the conclusion and future work are presented in Sect. 5.

2 Related Works

Several machine learning models (MLs) appeared in the educational field. First, we have searched for studies that have employed MLs in learning management systems to predict student performance.

In [7], the authors investigated the ability to predict the complexity of the coding, their approach is based on random forest (RF) to classify the complexity level of a software engineering team.

To improve learning in the educational field, the work in [8] provides a framework for managing the learning management system with the integration of data analysis and ML.

In the same context, there are many pieces of research that take into consideration the early prediction of performance to give a higher priority to learners that need guidance and help:

Using two models, logistic regression to classify pass/fail and linear regression to predict the final grades; the authors in [9] predict the performance in-between assignments for 4989 learners in 17 blended courses for the first ten weeks, then they remarked that as the number of weeks increases, so does the accuracy.

We can clearly see, in [10] studied the detection in the early stages of the performance (low or high) based on several methods, also give a full view of each stage along with methods that have high accuracy in the specific stage.

In other works, we find research that studied the behavior of learners. For example, authors of [11] used the Code Board Integrated Development Environment (IDE). They examined the learner's participation and behavior in Massive Open Online Courses (MOOCs) for programming. On the other hand, authors of [12] investigated how courses and a Learning Management System affected how students behaved and learned. We also see that in [13], instructional designs were examined to identify best practices while researching the educational scalability of MOOCs.

3 Dataset

Quantitative simulation research methodologies were used to achieve the objectives of the study according to the recommendations made in the framework phases illustrated in Fig. 1. To extract the most highly linked indicators, the data set will be prepared in these steps for visualization and clustering approaches, such as heat maps and hierarchical clustering. The best model will then be selected for forecasting the grade of final projects and all courses after the indicators have been employed in various classification algorithms.

The first step in our framework is preprocessing (cleaning, feature encoding, etc.); after that, we go through attribute selection (it is the phase where we train our algorithms to adjust parameters and find correlations). Finally, we analyze the models evaluated. More details will have birth in the next sub-sections.



Fig. 1. Analysis framework for datasets

3.1 Dataset Description

To address the study questions, we will work on a master class data gathered by [14] and divide it into two data.

- The first data have 270 instances (each instance is a course taken by a learner) and the following attributes:
 - student’s ID.
 - Age.
 - Bachelor’s degree name (Bsc deg).
 - Cumulative grade for that degree (Bsc GPA).
 - All courses that a student takes (Course).
 - Name of the instructor (instructor).
 - Course grade (grade).
 The target variable is the grade with (median = 4, mean = 3.3, and mode = 4).
- For the second data we have 38 instances and all the attributes of the previous one (are now dependent variables) plus the final project grade (target).

3.2 Pre-processing

In this stage, we try to remove all unimportant attributes to our goals (Course Id, Course Description, Academic year, etc.), and incomplete information (instances that have a lot of missing values, such as those who did not have grade values for the majority of their courses) was deleted. After cleaning the data and handling missing values, we have to encode the records; this is when we convert the nominal types (string format) into numerical type (integer) due to a variety of reasons, some ML techniques are effective and get the best accuracy with this size of data, such as the Multiple Perceptron Artificial Neural Network (MLP-NN) [15], Linear Discriminant Analysis (LDA) [16], Artificial neural networks [17], and also the Support Vector Machine algorithm (SVM), request for numerical types of attributes [18]. To transform our features to numerical type, for all these reasons, we encoded our features like this:

- Instructors: We have 4 instructors, so each one of them has a number from 1 to 4.
- Grades: We have 4 grades "A", "B", "C" and "Fail", respectively; we give them 4, 3, 2, and 1.
- Courses: We encoded them from 0 to 11, and 11 corresponds to the dissertation project.
- BSc Deg: We regroup all similar bachelors in the same category based on specialties.

In Tables 1 and 2, we give examples of our datasets and how they are organized like:

- In Table 1, we have 273 lines (instances) and 7 columns; the first is the Id of the students, from 2 to 6 are the attributes (depending variables) and the last is the target variable.
- For Table 2, we have 38 instances and 23 columns; the first is the student ID, from 2 to 22 are the attributes, and the last is the target (dissertation grade).

We remark that the outputs of Table 1 depend on Table 2.

As we mentioned earlier in this paper, the Attribute Selection phase is a part of the training stage of MLs algorithms because when the algorithm tries to adjust its parameters it is based on the relations between attributes.

Table 1. Dataset 1 example.

Student Id	Bsc deg.	Bsc GPA	Age	Course	Instructor	Grade
31	1	3	35	3	3	4
–	–	–	–	1	2	3
–	–	–	–	2	2	4
–	–	–	–	5	4	3
–	–	–	–	6	4	4
–	–	–	–	8	2	4
–	–	–	–	9	3	2
–	–	–	–	10	3	4

Table 2. Dataset 2 example.

Stu. ID	BSc Deg.	BSc GPA	Stu. Age	Course 1	Instructor 1	Grade 1
31	1	3	35	3	3	4
Course 2	Instructor 2	Grade 2	...	...	...	D. Grade
2	4	2	...	...	...	4

4 Results and Discussion

Since accuracy will be the primary critical statistic used to evaluate machine learning models, we will compare the accuracies of different machine learning algorithms most used in the literature such as Support Vector Machines (SVM), Random Forest (RF), Decision Tree (DT) and k-Nearest Neighbours (KNN) for both data sets 1 and 2.

Before citing our results we will discuss the best solution in our case from the view of the literature:

The authors of [8] conclude that naive-Bayes (NB) method gives the best result to predict student performance to reduce the number of drop-out students with a number of instances equal to 51.

But in their study [10], they tried to predict student performance early in the Learning Management System using several methods (NB, DT, logistic regression, MLP and SVM) in the entries 8,540,418, and their results show that DT and Multilayer perceptron (MLP) give the best accuracies.

In Fig. 2, we compare the algorithms most widely used in the literature in our case study to determine the best fit for this work.

To answer the second question of this research, we studied the main attributes responsible for the accuracy, we need to find the most correlated with the target variable:

The ability to forecast learners' grades, detect at-risk ones, and enhance system recommendations is a source of power to any educational institution.

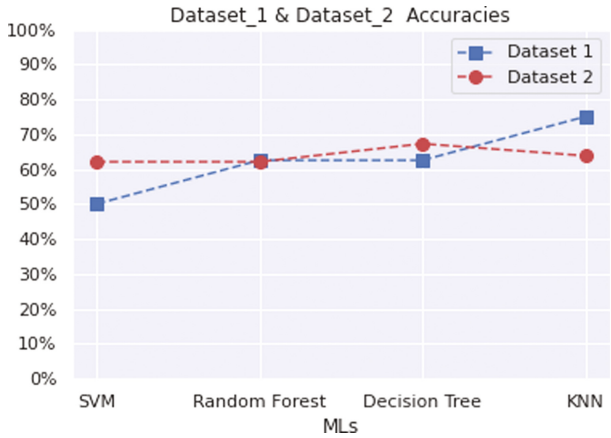


Fig. 2. Machine Learning algorithms comparison

Especially those who want to improve learners’ chances to cross the limit and be contributors, and learn from their educative interactions and experiences.

All institutions and universities want to have a low drop rate, so Predicting remarkable at-risk learners is thus a pressing need. Because by having the early information we can help them by easing the difficulties they are facing and recommending better experiences. Prior knowledge of students’ performance in each subject is also crucial.

Table 3. Accuracy for dataset 1.

Classification algorithms	Accuracy	Notes
SVM	50%	Using Gaussian kernel
RF (DV)	62.5%	With number of estimators: 120
DT	62.5%	With max-depth = 5
KNN	76%	At $K = 8$

However, making such forecasts is difficult because there are not enough dataset records to evaluate. However, our findings demonstrate that it is possible to do so with accuracy rates that are at least moderately significant. In comparison to other classifiers, the K-Nearest Neighbours classifier with $k = 8$ demonstrated the best ability to accurately predict student’s success in all courses (Table 3), and for the grade they would receive for their dissertation projects, we find the Decision Tree classifier with max-depth of 2 (Table 4).

Table 4. Accuracy for dataset 2.

Classification algorithms	Accuracy	Notes
SVM	62%	Using Gaussian kernel
RF (DV)	62%	With number of estimators: 40
DT	68%	With max-depth = 2
KNN	64%	At $K = 5$

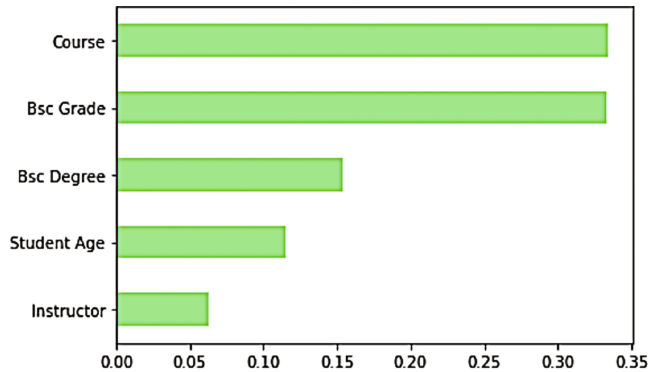


Fig. 3. Dataset 1 main attributes.

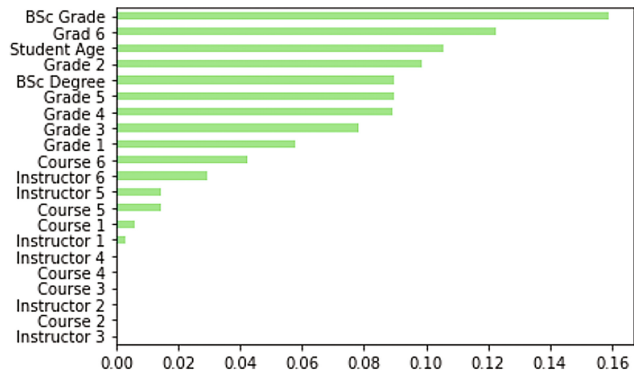


Fig. 4. Dataset 2 main attributes.

In general, there was a correlation between the grades and the final project, but the important key factors are the grade of course 6 and 2, age and Bsc. Grade (Fig. 4). On the other hand, when we look at (Fig. 3), we can clearly see that the course name and Bsc. Grades are the most correlated attributes to predict course grades.

5 Conclusion

Lately, the educational system has the technologies needed to cover more genres and levels of students and give the best learning experiences. However, this work enables machine learning to give a dashboard for educational institutions. Furthermore, it enables calendars management by generating classification models to detect at-risk students in order to help and guide them, as well as to find the key indicators for predicting grades and dissertation grades. Additionally, we evaluated some machine learning classification models to select the model that has the highest accuracy on our case study (small amount of data) for classifying final projects grade. The results obtained showed that Decision Tree and learning K-Nearest Neighbors algorithms deliver acceptable classification accuracy and reliability test rates.

For future work, we can work on hybridizing MLs algorithms to get better accuracy with a moderate time of execution. Also, we can investigate the early detection of students' classes.

References

1. Manjarres, A.V., Sandoval, L.G.M., Suárez, M.S.: Data mining techniques applied in educational environments: literature review. *Digital Educ. Rev.* (33), 235–266 (2018)
2. Martin, F., Chen, Y., Moore, R.L., Westine, C.D.: Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. *Educ. Technol. Res. Dev.* **68**(4), 1903–1929 (2020)
3. Comendador, B.E.V., Rabago, L.W., Tanguilig, B.T.: An educational model based on Knowledge Discovery in Databases (KDD) to predict learner's behavior using classification techniques. In: 2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), pp. 1–6. IEEE(2016)
4. Araque, F., Roldán, C., Salguero, A.: Factors influencing university drop out rates. *Comput. Educ.* **53**(3), 563–574 (2009)
5. Mueen, A., Zafar, B., Manzoor, U.: Modeling and predicting students' academic performance using data mining techniques. *Int. J. Modern Educ. Comput. Sci.* **8**(11), 36 (2016)
6. Asif, R., Merceron, A., Ali, S.A., Haider, N.G.: Analyzing undergraduate students' performance using educational data mining. *Comput. Educ.* **113**, 177–194 (2017)
7. Naseer, M., Zhang, W., Zhu, W.: Prediction of coding intricacy in a software engineering team through machine learning to ensure cooperative learning and sustainable education. *Sustainability* **12**(21), 8986 (2020)
8. Villegas-Ch, W., Román-Cañizares, M., Palacios-Pacheco, X.: Improvement of an online education model with the integration of machine learning and data analysis in an LMS. *Appl. Sci.* **10**(15), 5371 (2020)
9. Conijn, R., Snijders, C., Kleingeld, A., Matzat, U.: Predicting student performance from LMS data: a comparison of 17 blended courses using Moodle LMS. *IEEE Trans. Learn. Technol.* **10**(1), 17–29 (2016)
10. Riestra-González, M., del Puerto Paule-Ruíz, M., Ortin, F.: Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Comput. Educ.* **163**, 104108 (2021)

11. Gallego-Romero, J.M., Alario-Hoyos, C., Estévez-Ayres, I., Delgado Kloos, C.: Analyzing learners' engagement and behavior in MOOCs on programming with the Codeboard IDE. *Educ. Technol. Res. Dev.* **68**(5), 2505–2528 (2020)
12. Demmans Epp, C., Phirangee, K., Hewitt, J., Perfetti, C.A.: Learning management system and course influences on student actions and learning experiences. *Educ. Technol. Res. Dev.* **68**(6), 3263–3297 (2020)
13. Julia, K., Marco, K.: Educational scalability in MOOCs: analysing instructional designs to find best practices. *Comput. Educ.* **161**, 104054 (2021)
14. Zohair, A., Mahmoud, L.: Prediction of Student's performance by modelling small dataset size. *Int. J. Educ. Technol. High. Educ.* **16**(1), 1–18 (2019)
15. Ingrassia, S., Morlini, I.: Neural network modeling for small datasets. *Technometrics* **47**(3), 297–311 (2005)
16. Sharma, A., Paliwal, K.K.: Linear discriminant analysis for the small sample size problem: an overview. *Int. J. Mach. Learn. Cybern.* **6**(3), 443–454 (2015)
17. Pasini, A.: Artificial neural networks for small dataset analysis. *J. Thoracic Dis.* **7**(5), 953 (2015)
18. Naicker, N., Adeliyi, T., Wing, J.: Linear support vector machines for prediction of student performance in school-based education. *Math. Prob. Eng.* **2020** (2020)



Modeling of Psuedomorphic High Electron Mobility Transistor Using Artificial Neural Network

Radwa Mohamed¹(✉), Ahmed Magdy², and Sherif F. Nafea²

¹ Egyptian Customs Authority, Cairo, Egypt

Radwa.mohamed.pgs@eng.suez.edu.eg

² Faculty of Engineering, Suez Canal University, Suez, Egypt

{ahmed.magdi, sheriff_kamel}@eng.suez.edu.eg

Abstract. Modeling electronic devices has become a numerous work needed to keep up the development of the applications of the communication systems. Since the physical modeling of electronic devices is complex and time consuming. Recent research is using machine learning in modeling electronic devices, especially Artificial Neural Network (ANN) to reduce complexity and time consumption. In this paper, an enhanced model using ANN is proposed to model the scattering (S-) parameters in frequency range from 0.5 to 18 GHZ of GaAs Psuedomorphic High-Electron-Mobility-Transistor (pHEMTs). The proposed model uses four parameters which are the drain-source voltage, the drain-source current, the channel width, and the operating frequency as input parameters for the proposed ANN model to produce s-parameters for the pHEMT. Excellent agreements between the manufacturer's datasheet (ATF-34143) and model-calculated data are achieved. The ANN outputs are represented in the form of amplitude-phase and achieved higher performance than existing models. In modeling S_{11} the best training performance reached is $5.642 * 10^{-16}$ at epoch 392, in modeling at S_{21} the best performance reached is $2.869 * 10^{-21}$ at epoch 60, in modeling S_{12} the best performance reached is $4.3601 * 10^{-20}$ at epoch 20 and in modeling S_{22} the best performance reached is $5.7795 * 10^{-23}$ at epoch 144.

Keywords: Artificial neural network · s-parameters · Pseudomorphic high-electron-mobility-transistor

1 Introduction

In the last few years, due to the communications revolution, the electronic applications on these communication systems are varied like amplifiers, LNB(Low Noise Block), rectifications, antennas, etc. All these applications need many electronic devices, each one needs from the researchers to study its characteristics and behavior to determine its applications.

A high-performance type of field effect transistor known as a MESFET (Metal-Semiconductor Field-Effect Transistor) is primarily employed in demanding microwave

applications as a low noise signal amplifier and in higher power RF circuits. The junction FET (Field-Effect Transistor), sometimes known as a JFET, and the MESFET structure are quite similar (Junction Field Effect Transistor). The Schottky barrier diode junction is created by the MESFET's direct metal contact with silicon, as its name suggests. As a result, the Schottky diode functions in a similar manner to a JFET as a reverse biased diode. The main distinction is that the Schottky diode produces a much smaller diode. Silicon or another type of semiconductor material may be employed. Gallium arsenide GaAs, on the other hand, is the substance that is most frequently used. Gallium arsenide is typically chosen because it has extremely improved electron mobility, allowing for superior high frequency functioning [1].

PHEMT is a Monothetic Microwave Integrated Circuit (MMIC) technique that is used to create and produce integrated circuits. Due to its extremely wide-band performance qualities, such as low noise and great durability up to 40 GHz, pHEMT has become more widely used in MMICs such as Mini-Circuits. To achieve high-frequency performance, pHEMTs use heterojunctions between semiconductors with various compositions and bandgaps [2].

Researchers worked to study the characteristics of this type of devices. They studied the I-V characteristic and improved its accuracy and time [3] to build a reliable ML model and verify it using physical model. This paper study the s-parameters in GaAs pHEMT with amplitude-phase form to get high accuracy, although the complex number is more accurate in InP HBT [4]. The tried algorithms were chosen to achieve the best accuracy and time [5].

A comparison of the application of the ANNs for modelling the scattering (S) parameters of a variety of FET technologies versus bias point, ambient temperature, and geometrical dimensions is conducted in this review of the (FET) small-signal modelling utilizing (ANNs) [7].

Using nonlinear autoregressive series-parallel and parallel architectures to simulate a $2 \times 200 \mu\text{m}$ device for a wide frequency range of 1 GHz–18 GHz, tiny signal behaviour of GaN (HEMT) is being modelled. Based on the training procedure, accuracy, convergence rate, and number of epochs, a comparison of the two architectures. For the entire broad frequency range, there is excellent agreement between the suggested model and the measured S-parameters [9].

Neural network behavior modelling is utilized to simulate the large-signal nonlinear behavior of GaN P-HEMTs. When modelling drain current, output power, and gain, the impacts of Back Propagation BP neural network method, GA-BP neural network algorithm, and GA-ELM neural network algorithm are compared. The GA-ELM algorithm model's modelling effect is discovered to produce improved accuracy [2].

Therefore, the employment of ANN in the development of the pHEMT model is comprehensively investigated in this study. The ANN was chosen because it demonstrates certain traits like noise parameters (R parameters) and scattering parameters (S parameters). Additionally, the ANN runs at a high speed and only needs the dataset to be quickly and easily prepared. The k-fold cross-validation method is also used by the ANN suggested model to prepare the dataset since it is simple to apply and produces skill estimates that are typically less biased than those produced by other techniques. The drain-source voltage (VDS), drain source current (IDS), operating frequency (F), device

width (w), and outputs of scattering parameter amplitude (S “amplitude” and scattering parameter phase (S “angle”) make up the proposed ANN model for the (ATF-34143) PHEMT.

The subsections are organized as follows. Section 2 displays the problem formulation information. Section 3 discusses the theory of the proposed model. Section 4 provides the simulation results and assessment of the proposed approach. The Sect. 5 concludes the paper.

2 Problem Formulation

A high dynamic range, low noise PHEMT is the (ATF-3414). It boasts an 800-micron gate width, good homogeneity, and affordable surface mount. Due to its exceptional low noise figure and high linearity combination, it is perfect for the first stage of base station LNA (Low Noise Amplifier). The device is also appropriate for wireless LAN, WLL (Wireless Local Loop)/RLL (Radio Local Loop), MMDS (Microwave Multipoint Distribution Service), and other systems requiring extremely low noise figures in the 450 MHz to 10 GHZ frequency range.

The manufacturer’s datasheet [12] is where the data came from. The predictors have the following values: VDS = 3 V, IDS = 20 mA, W = 800 m, and a frequency range of 0.5 to 18 GHz. S-parameters for magnitude and phase are the outputs.

The whole diagram of the models is shown in Fig. 1. It would be beneficial to design the preprocessing processes before training the model because preprocessing has been found to have a direct impact on the computing efficiency of the algorithms, which in turn affects the model’s correctness. The common scale is achieved by dividing each column by its corresponding maximum value. These generated variables with a 0–1 range. The model’s performance is then assessed by examining the MSE obtained using (1). MSE = mean square error, n = number of data points, y_i = observed values, $\tilde{y}_i$ = predicted values.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \tilde{y}_i)^2 \quad (1)$$

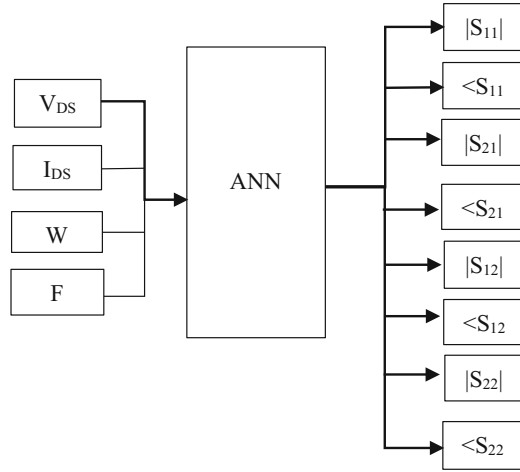


Fig. 1. A block diagram for the proposed ANN model

3 System Model

The models are developed using MATLAB due to its easy-to-understand syntax and array of ML-specific tools and frameworks. The input predictors are V_{DS} , I_{DS} , W , and F , as was previously mentioned. The scattering parameters are the outputs. The flowchart in Fig. 2 describes the steps taken both before and after training the model. The import of the data is followed by the selection of the predictors and predicated variables. The procedure outlined in Sect. 2 is used for preprocessing.

Data is split into 80% for training and 20% for testing in order to build the training and testing sets. The employed training function is (Levenberg-Marquardt). The most important part of the machine learning algorithm is fine-tuning the parameters. Fivefold cross-validation is used for tuning. Three layers [4-8-16-2-2] with an epoch of 5000 make up the forward ANN.

Once the ideal parameters have been determined, the ANN model is constructed with the ideal parameters, and the model’s performance is determined using MSE. The curves depicting the measured and simulated s-parameters are drawn in Figs. 7,8,9,10, which demonstrate a good agreement, allowing for a closer examination of the model’s effectiveness.

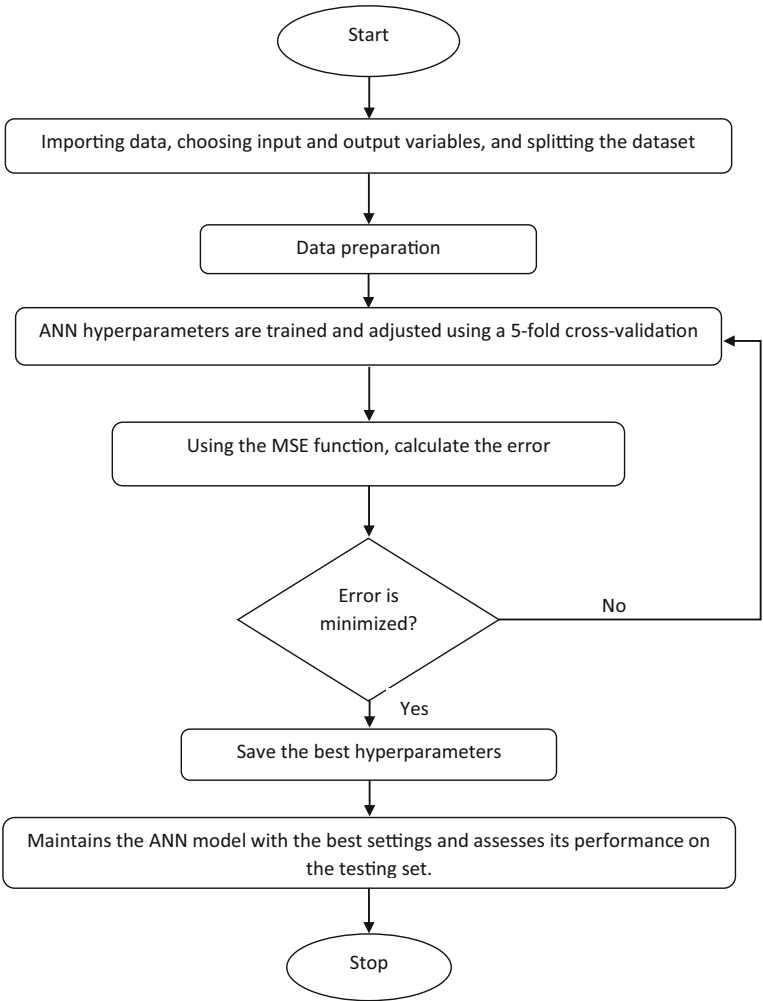


Fig. 2. Flowchart shows the actions taken to train and test the ANN model

DataSet

Data taken from the device’s manufacturer’s datasheet (ATF-34143) [12] were VDS = 3 V, IDS = 20 mA, W = 800 m, and a frequency range of 0.5 to 18 GHz. There are

two outputs: S (amplitude) and S . (phase). 80% of the data is used to train the ANN, and 20% is used to test the ANN.

4 Simulation Results and Discussion

- The proposed model consists of three-hidden layers with 8 nodes in first layer, 16 nodes in second layer and 2 nodes in third layer.
- The ANN model uses training function Levenberg-Marquardt (LM) forward propagation algorithm (FP) with epoch = 5000. The MSE is checked the performance of the modelled scattering parameters S_{11} , S_{21} , S_{12} , S_{22} and the outcome is shown in Fig. 3, Fig. 4, Fig. 5 and Fig. 6 respectively.
- It is shown that in modeling S_{11} the best training performance reached is 5.642×10^{-16} at epoch 392, in modeling at S_{21} the best performance reached is 2.869×10^{-21} at epoch 60, in modeling S_{12} the best performance reached is 4.3601×10^{-20} at epoch 20 and in modeling S_{22} the best performance reached is 5.7795×10^{-23} at epoch 144.

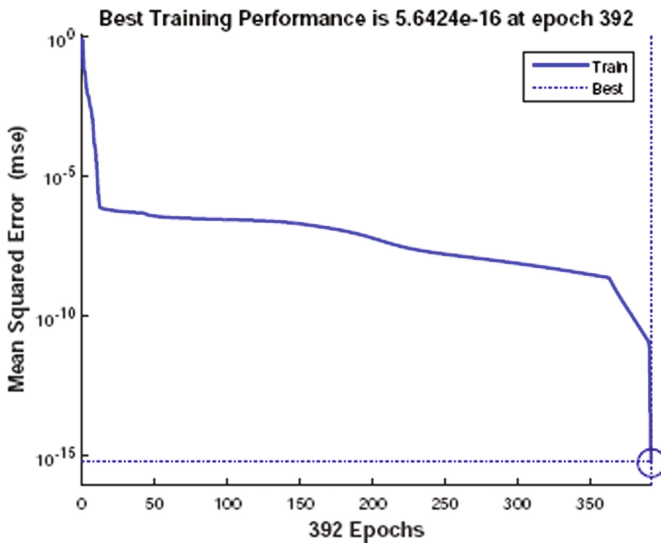


Fig. 3. S_{11} training performance

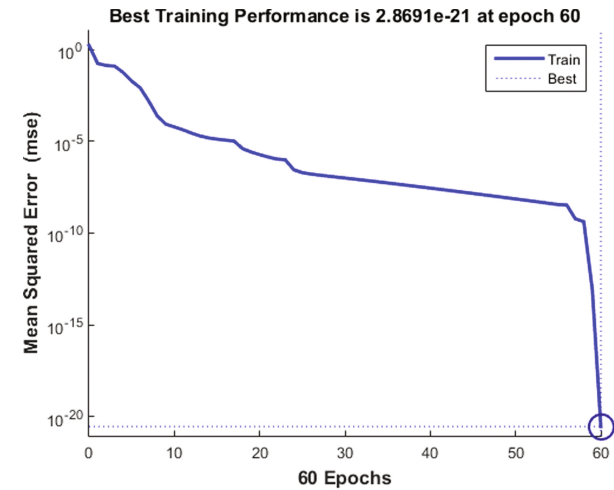


Fig. 4. S_{21} training performance

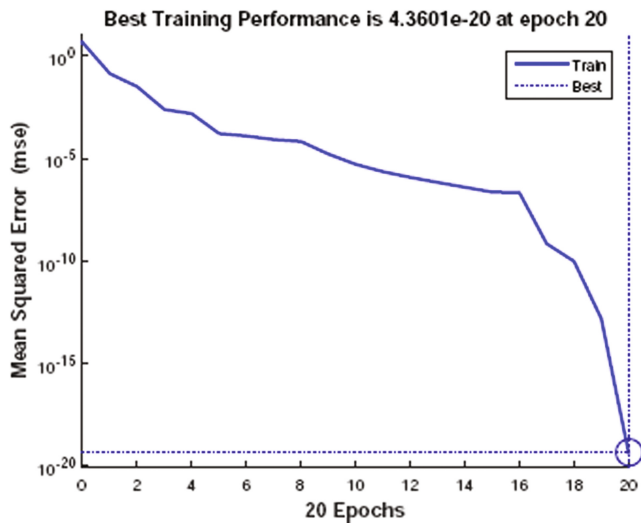


Fig. 5. S_{12} training performance

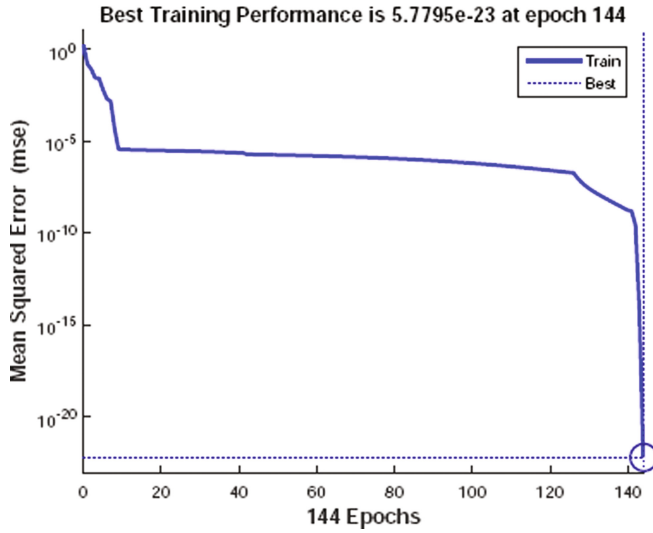


Fig. 6. S_{22} training performance

Additionally, Figs. 7, 8, 9, and 10 show, respectively, the observed and simulated plots for the four scattering parameters S_{11} , S_{21} , S_{12} , and S_{22} for a broad range of frequencies from 0.5 to 18 GHz.

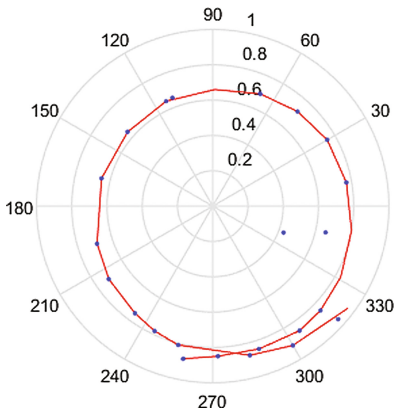


Fig. 7. Comparison of the modeled (symbols) and simulated (lines) S_{11}

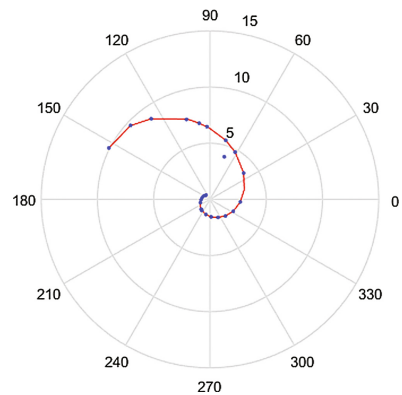


Fig. 8. Comparison of the modeled (symbols) and simulated (lines) S_{21}

The ANN provides excellent fitting for the measurements. The same results are obtained with multiple running of the program, and this validates the accuracy of the solution presented by FP based ANN.

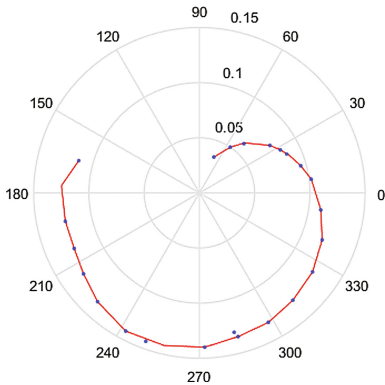


Fig. 9. Comparison of the modeled (symbols) and simulated (lines) S_{12}

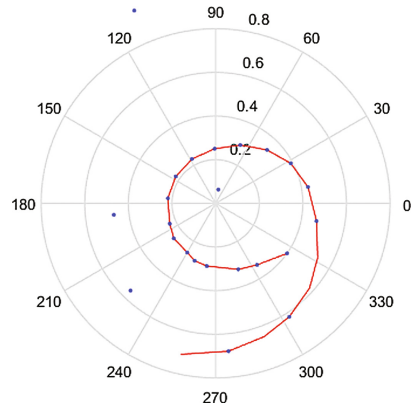


Fig. 10. Comparison of the modeled (symbols) and simulated (lines) S_{22}

5 Conclusion

The ANN can be successfully applied to model the scattering parameters of GaAs pHEMT. This paper proposed an enhanced ANN model for GaAs pHEMT with high accuracy using multilayer ANN. The inputs are $W = 800 \mu\text{m}$, $V_{DS} = 3 \text{ V}$, $I_{DS} = 20 \text{ mA}$ and frequency of range 0.5–18 GHz. The ANN has two outputs S (amplitude) and S (phase). The developed model is compared in term of MSE. In modeling S_{11} the best training performance reached is 5.642×10^{-16} at epoch 392, in modeling at S_{21} the best performance reached is 2.869×10^{-21} at epoch 60, in modeling S_{12} the best performance reached is 4.3601×10^{-20} at epoch 20 and in modeling S_{22} the best performance reached is 5.7795×10^{-23} at epoch 144. The ANN model offers simple formulation and can be ready to adapt in CAD tools.

References

1. Camarchia, V., Cappelluti, F., Ghione, G., Limiti, E., Moran, D.A.J., Pirola, M.: An overview on recent developments in RF and microwave power H-terminated diamond MES-FET technology. In: 2014 International Workshop on Integrated Nonlinear Microwave and Millimetre-wave Circuits (INMMiC), pp. 1–6. IEEE (2014)
2. Wang, S., Zhang, J., Liu, M., Liu, B., Wang, J., Yang, S.: Large-signal behavior modeling of GaN P-HEMT based on GA-ELM neural network. *Circ. Syst. Signal Process.* **41**(4), 1834–1847 (2021). <https://doi.org/10.1007/s00034-021-01891-7>
3. Hari, N., Ahsan, M., Sridhar, R., Padmanaban, S., Albarbar, A., Blaabjerg, F.: Gallium nitride power electronic devices modelling using machine learning. *IEEE Access* **8**, 119654–119667 (2020)
4. Perkins, N., Abrokwha, J., Rohdin, H.G., Marsh, P., Stanback, J.: Avago technologies general IP Singapore Pte Ltd. Pseudomorphic high electron mobility transistor (pHEMT) comprising low temperature buffer layer. U.S. Patent 8,901,606 (2014)

5. Zhang, A., Gao, J.: InP HBT small signal modeling based on artificial neural network for millimeter-wave application. In: Proceedings of the 2020 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization (NEMO), Hangzhou, China, 7–9 December 2020, pp. 1–3 (2020)
6. Deng, W., Zhang, W., Peng, Y., Wu, W., Huang, J., Luo, Z.: Compact modeling of metal-oxide TFTs based on artificial neural network and improved particle swarm optimization. *J. Comput. Electron.* **20**(2), 1043–1049 (2021). <https://doi.org/10.1007/s10825-020-01641-z>
7. Marinkovic, Z., Crupi, G., Caddemi, A., Markovic, V., Schreurs, D.M.M.-P.: A review on the artificial neural network applications for small-signal modeling of microwave FETs. *Int. J. Numer. Model.* **33**(3), e2668 (2020)
8. Qi, J., et al.: Small-signal modeling of microwave transistors using radial basis function artificial neural network-comparison of different methods for spread constant determined. Wiley (2022)
9. Khusro, A., Husain, S., Hashmi, M.S., Auyuneur, M., Ansari, A.Q.: A reliable and fast ANN based behavioral modeling approach for GaN HEMT. In: 2019 16th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD), pp. 277–280. IEEE (2019)
10. Zhu, Y.Y., Ma, J.G., Fu, H.P., Zhang, Q.J., Cheng, Q.F., Lin, Q.: Accurate modeling of pHEMT output current derivatives over a wide temperature range. *Int. J. Numer. Model. Electron. Netw. Dev. Field.* **30**(3–4), e2185 (2017)
11. Husain, S., Begaliyeva, K., Aitbayev, A., Chaudhary, M.A., Hashmi, M.: Decision tree based small-signal modelling of GaN HEMT and CAD implementation. In: 2022 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–6. IEEE (2022)
12. https://www.alldatasheet.com/view.jsp?Searchword=Atf-34143%20datasheet&gclid=EAIaIQobChMI9PLYps-0-gIVx-Z3Ch1aNQndEAAYASAAEgJEYvD_BwE. Accessed Nov 2022
13. Hari, N., Chatterjee, S., Iyer, A.: Gallium nitride power device modeling using deep feed forward neural networks. In: 2018 1st Workshop on Wide Bandgap Power Devices and Applications in Asia (WiPDA Asia), pp. 164–168. IEEE (2018)
14. Wei, J.H., Mao, W., Fang, H., Zhang, Z., Zhang, J.X., Lan, B.J., Wan, J.: Advanced MOS-FET model based on artificial neural network. In: 2020 China Semiconductor Technology International Conference (CSTIC), pp. 1–3. IEEE (2020)



Leveraging Blockchain and Machine Learning to Improve IoT Security for Smart Cities

Mayar M. Moawad^(✉) , Magda M. Madbouly , and Shawkat K. Guirguis 

Department of Information Technology, Institute of Graduate Studies and Research,
Alexandria University, 163 Horreya Avenue, El-Shatby, P.O. Box 832, Alexandria 21526, Egypt
{igsr.mayar.mostafa, mmadbouly, shawkat_g}@alexu.edu.eg

Abstract. Recently, Internet of Things (IoT) infrastructures are developing various applications in sustainable smart cities and societies. However, there are numerous challenges in smart cities, such as security, privacy, trust, verifiability, communication latency, scalability, and centralization preventing faster adaptations of IoT. Machine Learning (ML) is an important analytic tool that provides a scalable and accurate analysis of data in real time. However, there are several obstacles to designing and developing a usable large data analysis tool utilizing ML, such as centralized architecture, security, and privacy, resource limits, and a lack of sufficient training data. Blockchain, as opposed to that, promotes a decentralized architecture as new technology. It encourages the secure sharing of data, and resources among the various nodes of the IoT network, removing centralized control and overcoming ML's current difficulties. As a result, this study provides a smart city intrusion detection system. This system consists of three modules: a trust module based on designing an address-based blockchain reputation system, a two-level privacy module based on blockchain-based enhanced Proof of Work technique, and an intrusion detection module. We provide a blockchain-IPFS integrated Edge-Fog-Cloud infrastructure, named Cloud-Block, Fog-Block, and Edge-Block, to utilize the system proposed for smart cities, related to the inherited strengths and shortcomings of Edge-Fog-Cloud architecture. The sBoT-IoT and TON-IoT datasets are utilized to evaluate the system. Finally, a comparison of our implementation results shows that our system outperforms other state-of-the-art systems.

Keywords: Blockchain · Cyber security · Internet of Things (IoT) · Intrusion detection system · ML · Smart cities

1 Introduction

The smart city refers to the emerging Information and Communication Technology industries in addition to the operations and environment of urban areas [1]. There is a global trend toward smart cities as global urbanization continues to rise, with the overall population predicted to double by 2050. Therefore, the approach of smart cities has emerged as the solution to address the aforementioned challenges [2]. Technological advancement and rapid advancement of low-cost devices such as actuators, sensors, and

Internet of Things (IoT)-oriented infrastructure, have become a crucial environment in developing various smart cities applications [3]. IoT is a network of smart devices that exchange data over the internet. Smart devices are utilized to acquire data in a different environment, and some events are triggered as a result. A wide range of effectual services in the smart city as smart transportation, smart waste collection, smart healthcare, smart building, smart homes, smart power systems, etc. [4–6]. It's among the most emerging technologies in the last decade and has uses in numerous application areas. Security and privacy are still challenges in many applications area. Some research on IoT security and privacy issues has already been completed [7].

Artificial Intelligence (AI) has grown in popularity in recent years as a consequence of advances in Machine Learning (ML), particularly Deep Learning, (DL) and Reinforcement Learning (RL), demonstrating its utility in a wide range of applications where classification or regression problems play an essential part [8]. ML is one of the AI techniques which trains machines using different algorithms and helps devices learn from their experience instead of programming them explicitly [9]. ML does not require human intervention or rigorous mathematical formulas to function in dynamic networks. ML methods have advanced significantly in recent years for IoT security purposes. [10]. Therefore, ML techniques should be utilized to detect various IoT attacks in their early stages by analyzing device behavior. In addition, appropriate solutions can be provided using different ML algorithms for resource-limited IoT devices.

ML for IoT security is divided into the following two categories for IoT devices i.e., to detect smart attacks and establish a strong defensive policy, ML techniques such as supervised techniques, unsupervised techniques, and reinforcement learning can be utilized [11] and ML-based solution for IoT security field has become an emerging research area and is attracting the attention of today's researchers to add more to this field over the last few years. These remedies have been explored based on three main architectural layers of an IoT system, including the physical/perception layer, network layer, and web/application layer-wise. Figure 1 a few of the current ML techniques in IoT security [11].

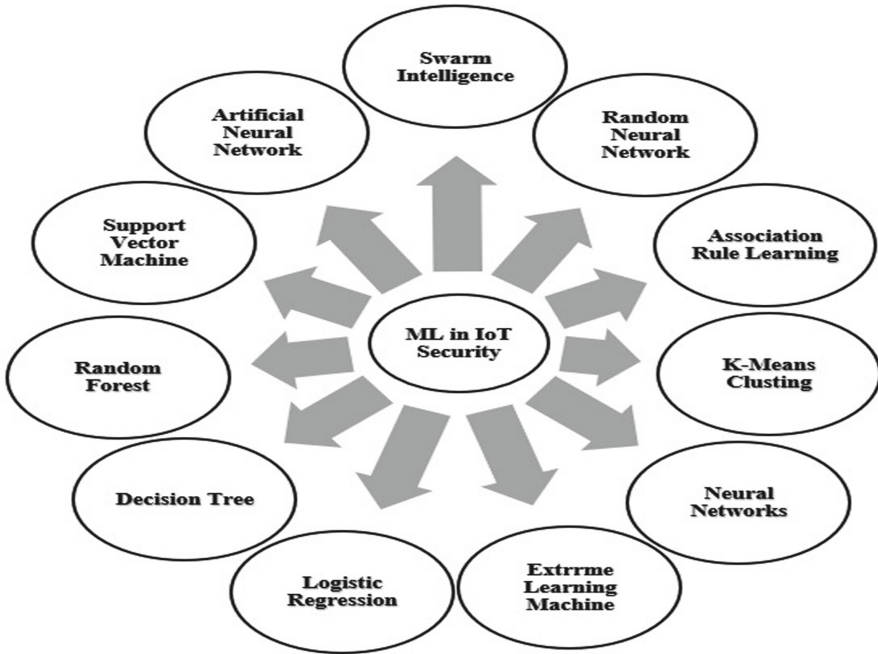


Fig. 1. Existing ML techniques in IoT security.

On the contrary, blockchain supporting a secure network over untrusted parties has recently become an attractive choice to give a secure, scalable IoT network [12]. It provides privacy in an IoT network by delivering the silent features of decentralization and an immutable distributed ledger. Decentralization in blockchain enables the secure sharing of resources, and data among all member nodes in an IoT network, overcoming many-to-one traffic flows and reliance on central control and providing enough data for big data analysis. As a result, it eliminates the problem of a single point of failure and insufficient training data, and it reduces communication delay. Furthermore, the immutable distributed ledger securely maintains information about all IoT network member nodes in the form of blocks cryptographically controlled by all member nodes. This feature prevents malicious tampering with training data, allowing an IoT network to respond very well to data poisoning attacks [13].

The paper proposes an enhanced Intrusion Detection IoT Security System to help solve Cyber security problems based on ML. To make the network more efficient and autonomous, ML techniques are being integrated into IoT. One type of ML is DL, which is computationally complex and expensive [8]. One of the challenges is combining DL methods with IoT to boost the effectiveness of IoT applications. Integration of these methods with an emphasis on balancing computational cost and efficiency is critical for next-generation IoT networks. Considering the requirements of DL, and IoT and seamless integration necessitates a complete overhaul of the communication stack from the physical layer to the application layer [8]. The ML task in IoT and Blockchain can solve IoT challenges.

The paper is structured as follows: the next section describes the related work in IoT, Sect. 3 discusses the experimental design of the system proposed, and Sect. 4 represents results & discussion of the system proposed. Finally, the conclusion and future work of the proposed study is in Sect. 5.

2 Literature Survey

Many researchers have studied and discussed unresolved research issues of Blockchain for AI, and AI for Blockchain on IoT. Atlam et al. [14] described a view of IoT integration, and AI, which has benefits and opportunities of AI in various IoT applications, and presented the challenges, and discussion for successful convergence of IoT with AI. Authors in [15] classified blockchain technology into 4 categories: (a) key characteristics, and taxonomy of blockchain, (b) algorithm of consensus, (c) blockchain applications, and (d) technical challenges in current research. Wright et al. [16] introduced an Ethereum-based smart edge for smart contract computing and demonstrated that it is a low-cost, high-accuracy resource management tool. It enables nodes to offload verifiable calculations to edge devices in exchange for payment. In [17] authors discussed the advantages of Blockchain thinking for AI and human enhancement, as well as an architectural proposal for Blockchain thinking and the processing of Blockchain thinking in AI. Authors in [18] presented a summary of Blockchain applications for AI and identified open research challenges of utilizing Blockchain for AI. They reviewed literature and summarized applications based on blockchain, also platform protocols targeting AI areas.

Qian et al. [19] presented three layers of IoT, perception, network, and application layers, as well as security issues in each layer, high-level security management methods depending on blockchain technology for various IoT devices, and feasible research problems. Their research included irregular traffic of the network monitoring and identity verification using ML. Rathore et al. Blockchain secure DL method is proposed in [13] to provide reliable data in IoT applications by bringing AI, and blockchain together at the device layer. It achieves high accuracy while also having a high latency for IoT data. Alguliyev et al. [20] proposed a privacy-preserving DL mechanism. For big data inquiry, this model employs a denoising autoencoder in conjunction with convolutional neural networks. Makhdoom et al. [21] recently proposed a Privy Sharing named framework. As stated in the previous section, existing research faces significant protests such as lack of privacy, low accuracy, centralization, low latency, and massive bulk of data. Our study differs significantly from other existing studies as it provides integration, extensiveness, and comprehensiveness to the convergence of DL, and blockchain for IoT-driven smart cities and their networks.

3 Proposed Methodology

3.1 System Information

The system is implemented using Python and R programming languages. Blockchain Technology is implemented using Ethereum, and Solidity programming language (version 6.0). IPFS version 0.4.19 has been configured in the system with the following

specification: Intel(R) Core (TM) i7-7500U CPU @ 2.70 GHz ($\times 64$ -based Processor), 8 GB RAM, and PC running Microsoft Windows 10 Pro Enterprise 64 bit.

3.2 Dataset Information

1. **BoT-IoT dataset** [22] is derived in the Canberra Cyber Range Lab, and a realistic network environment is designed. This data collection is generated using the Message Queuing Telemetry Transport protocol, which links machine-to-machine communications, making it a viable option for IoT solutions. The dataset contains various attacks such as DDoS, DoS, and Theft.
2. **TON-IoT dataset** [23] is derived from a practical and large-scale network created by the UNSW Canberra Cyber IoT Lab, School of Engineering and Information Technology (SEIT), UNSW Canberra. The TON-IoT datasets are the next generation of Industry 4.0/IoT and Industrial IoT (IIoT) datasets for checking the dependability and effectiveness of various AI-based cybersecurity applications. In parallel processing, the dataset is compiled to compile a variety of routine, and cyber attack events from IoT networks. A new testbed has been built at the IoT lab to connect a variety of virtual computers, physical tools, hacking systems, cloud, fog, and edge systems, and IoT sensors to simulate the functionality and scalability of automotive IoT and Enterprise 4.0 networks. This dataset contains a variety of recent smart city-based attacks, such as DoS, DDoS, and ransomware, that have been launched against web applications, IoT gateways, and computer systems throughout the IoT network.

3.3 Proposed System

A deployment architecture for intrusion detection systems in a smart city is presented. The system proposed is expanded on the edge side, fog side, and cloud side to address the shortcomings of standalone architectures and to take the privileges of collaborative networks. Furthermore, to reap the benefits of blockchain technology completely, we have integrated blockchain and IPFS at the edge, fog, and cloud sides. As illustrated in Fig. 2 the system proposed is deployed in blockchain and IPFS-enabled edge, fog, and cloud. The systematic deployment architecture at Edge-Block, Fog-Block, and Cloud-Block side, for the system proposed, contains three main components; blockchain-based trust management reputation system, privacy, and preservation module, and ML detection module as discussed below:

Once the data (request for a service) is generated from IoT devices at the device layer of the smart city. The network traffic is forwarded to Edge Block using the nearest gateway/router. At Edge Block, the former consists of a sensor (packet sniffing tool) to sniff the incoming traffic and extracts the associated features. To ensure the trust for IoT sensor nodes' data at the Device layer, an address-based blockchain reputation system is designed to compute the reputation score for each registered IoT node, based on transaction and confidence threshold value. A higher reputation score of a node contributes to higher trust in the true observation of the node using Algorithm 1. Depending on the outcome, the transaction is categorized into three types: Honest, General, and Dishonest. Using Algorithm 2, the transaction or raw data (with trust value) is saved

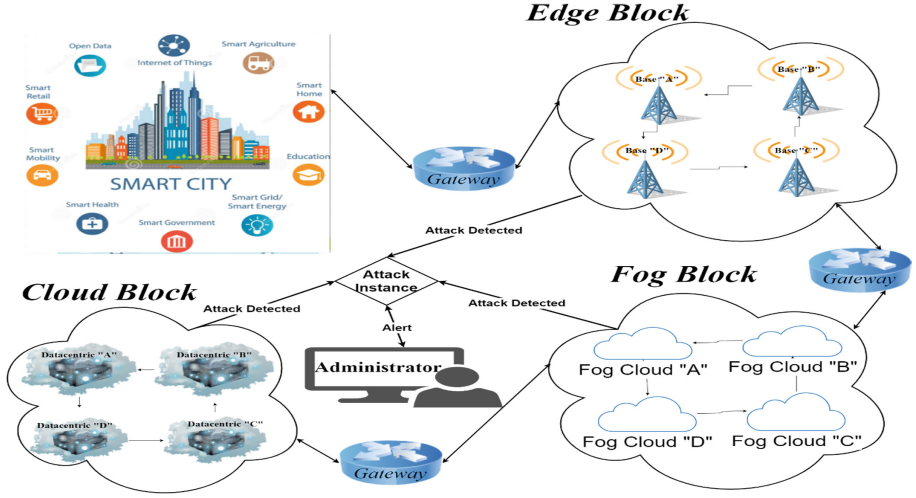


Fig. 2. Proposed system in edge block, fog block, and cloud block of a smart city.

in the distributed file storage system IPFS [5, 24], while in the blockchain network the message digest containing Reputation score (Rps) and Transaction score ($Txscore$) of 3 types for nodes are stored.

Algorithm 1
Calculating Transaction score & Reputation score
Input:
Transaction score ($Txscore$), Confidence Threshold (CTh), Dataset (D), Transaction value ($Txvalue$), Reputation score (Rps), Number of transactions (No_of_Tx), Number of feature count in Dataset ($f.count$)
<pre> 1. $Txscore=0$ 2. $CTh = \text{range}(\min(di), \max(di))$, where $di \in D$. 3. if ($Txvalue == CTh$) then 4. $Txscore += 1$ 5. else 6. $Txscore += 0$ 7. end if 8. $Rps = ((Txscore)/10) / No_of_Tx$ 9. if ($(Rps > (f.count - 4)/10)$ and ($Rps \leq (f.count)/10$)) then 10. print ("Honest Transaction"); 11. else if ($(Rps > (f.count - 8)/10)$ and ($Rps \leq (f.count - 4)/10$)) then 12. print ("General Transaction"); 13. else 14. print ("Dishonest Transactions"); 15. end if </pre>
Output:
The transaction is classified into three categories

Algorithm 2	
An IPFS mechanism for IoT-driven smart city	
1.	for All Sensor nodes (SN.z) do
2.	Retrieve the sensor node <i>info</i> (Txvalue)
3.	Tx = COMPUTE_REPUTATION(Txvalue)
4.	if ((Tx == ‘‘Honest Transaction’’) OR (Tx == ‘‘General Transaction’’) OR (Tx == ‘‘Dishonest Transaction’’)) then
5.	Sends this <i>info</i> to the IPFS node
6.	IPFS stores this <i>info</i> in the distributed hash table (DHT)
7.	IPFS assigns the unique addressable hash of the sensor <i>info</i>
8.	Send the addressable hash into Edge Block, Fog Block, and Cloud Block for further access to actual data
9.	end if
10.	end for

The raw data is sent to the privacy and preservation module along with the trust data (three categories). In this module, a message digest with proof of hash is generated using blockchain-based enhanced proof-of-work (ePoW). The message digest is then distributed throughout the blockchain network. This method verifies chains of data records and prevents inference attacks that could learn from system-based ML. Finally, the DL algorithm [25] is utilized to categorize data into normal and different attack types, depending on which anomalous class administrator is alerted. Finally, for a normal transaction the required information (*info*), if available at the edge side is provided and in case the *info* is unavailable, the security *info* along with the request is sent to Fog Block. At the Fog Block, the set of fog nodes builds a blockchain ledger for on-chain storage and IPFS for off-chain storage.

Once the Fog Block receives a request for the *info*, the aforementioned procedure is followed and service is provided to the requester for attack instances administrator is alerted and in case of *info* is not available, the security *info* along with the request is sent to Cloud Block. Cloud Block consists of various types of data centers provided by various vendors. For the system proposed, we have utilized four data centers A, B, C, and D as shown in Fig. 1. These are the entities that comprise the Cloud-Block network; the system proposed is deployed at each data center to form a blockchain network. This approach builds user trust by making the network immutable, auditable, and verifiable [26]. When the Cloud Block receives a request for *info*, the aforementioned procedure is followed, and the service is given to the requester, while the administrator is notified of any attack instances. The system proposed takes the advantage of collaborative cooperation between edge-fog-cloud architecture and integrates Blockchain, and ML to scheme a sustainable smart city.

4 Experimental Results and Discussion

4.1 Description of Evaluation Metrics

The intrusion detection system’s performance most commonly rely on the evaluation metrics such as; Accuracy, Detection rate (DR), Precision, and F1-Score. To compute these scores, various parameters were used that are discussed below [27]:

- True Positive (TP): This determines the number (No.) of malicious vectors in IoT network traffic, that are correctly classified as anomalous behavior by the system.
- True Negative (TN): This determines No. of normal vectors in IoT network traffic, which is correctly classified as normal behavior by the system.
- False Positive (FP): This determines No. of normal vectors in IoT network traffic, which is incorrectly classified as anomalous behavior by the system.
- False Negative (FN): This determines No. of anomalous vectors in IoT network traffic, which is incorrectly classified as normal behavior by the system.

The evaluation metrics are calculated using the aforementioned parameters and are discussed below [5, 28]:

- (a) Accuracy (AC): It determines the ratio of correctly identified instances by the system to the total No. of observations present in the test collection. It takes both TP and TN into account for calculating the accuracy of the system.

$$AC = \frac{TP + TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N} \quad (1)$$

- (b) DR: It determines No. of malicious activities observed by the system, divided by the total No. of activities present in the test collection. It is also known as Recall (RC).

$$DR = \frac{TP}{TP + FN} = \frac{TP}{P} \quad (2)$$

- (c) Precision (PR): It determines No. of malicious activities observed, divided by the total No. of observations defined as an attack by the system.

$$PR = \frac{TP}{TP + FP} \quad (3)$$

- (d) F1 Score: It calculates the weighted mean of PR and DR. It is utilized when class distribution is imbalanced and is more useful than accuracy as it takes FP and FN into account for calculation.

$$F1\ Score = \frac{2 * PR * DR}{PR + DR} = \frac{2TP}{2TP + FP + FN} \quad (4)$$

4.2 Results and Discussion

Experiment 1: Measuring the Performance of Detection Rate with different Attacks Types While Implementing Proposed Architecture Facilitates the Detection of Attacks Effectively

The system's results are compared to some recent state-of-the-art. For evaluation, TP2sf [5] is used, and the conclusion of the preceding algorithm is compared to the suggested architecture using two IoT-based datasets. The system performance is checked using the class-specific DR, overall accuracy, PR, and F1-Score. Figure 3 and Fig. 4

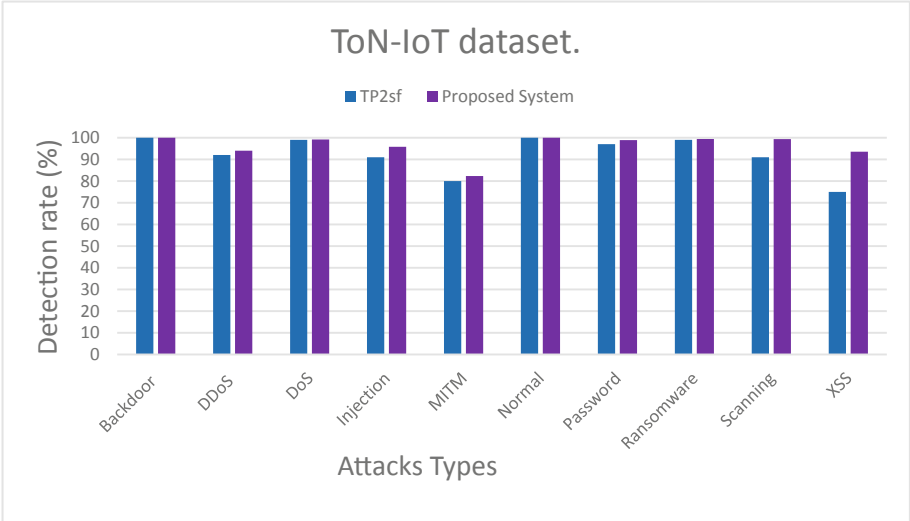


Fig. 3. Detection rate using TON-IoT dataset.

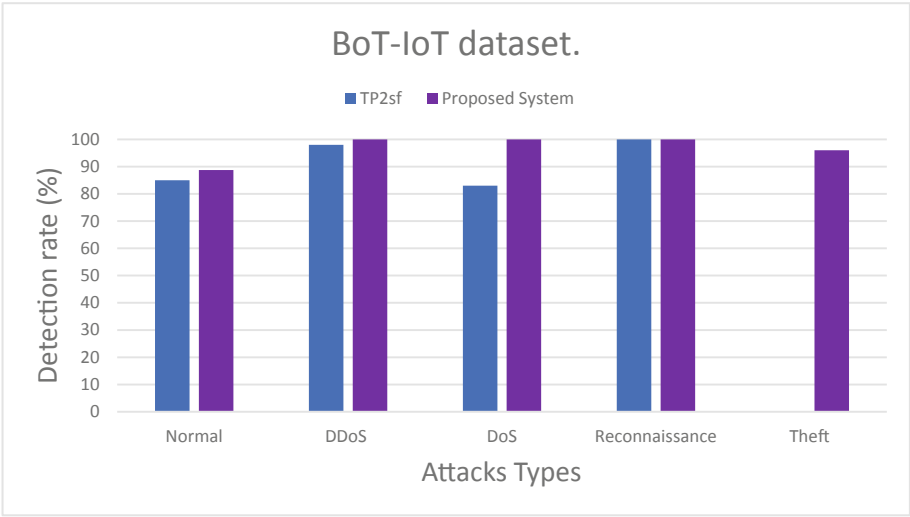


Fig. 4. Detection rate using BoT-IoT dataset.

show attack types using DR comparison with TP2sf on TON-IoT and BoT-IoT datasets respectively. It is recognized that; the suggested architecture outperforms in almost all cases for identifying various attack vectors present in both datasets.

Experiment 2: Comparative Study to Show that Implementing DL Facilitates the Detection of Attacks Effectively

Figure 5 and Fig. 6 illustrate the overall AC, PR, DR, and F1-Score comparison. It is worth noting that, using TP2sf for the measurement of TON-IoT, and BoT-IoT datasets. Therefore, it can be concluded that the architecture has shown promising results compared to another model. The previously discussed method has improved significantly the proposed architecture's overall performance. Furthermore, incorporating blockchain technology into Edge, Fog, and Cloud infrastructures improve verifiability, traceability, and reliability.

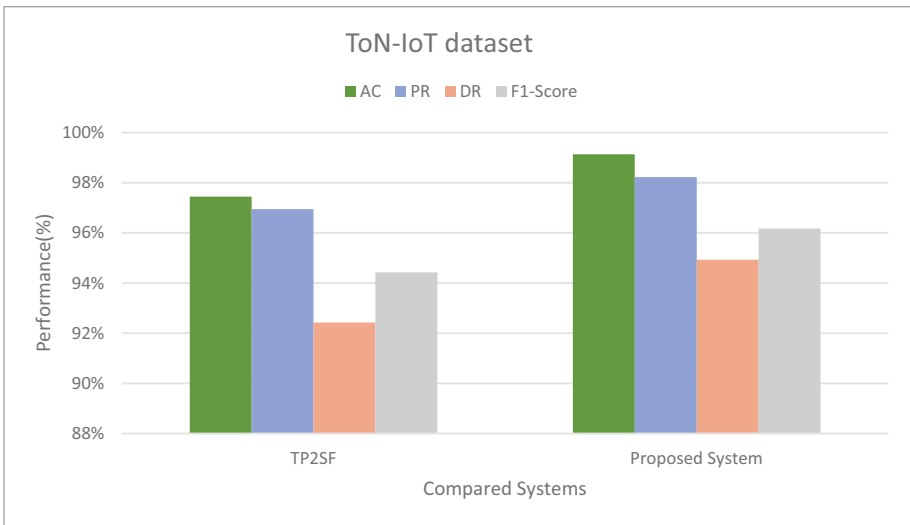


Fig. 5. Comparison in regards to AC, PR, DR, and F1-Score using the TON-IoT dataset.

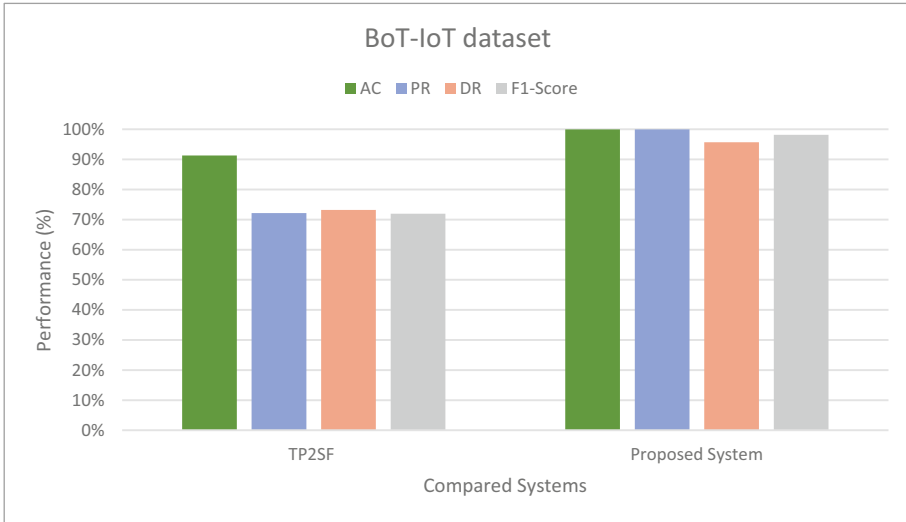


Fig. 6. Comparison in regards to AC, PR, DR, and F1-score using BoT-IoT dataset.

5 Conclusion

An Intrusion Detection System for improving a sustainable smart city utilizing blockchain and ML methods is introduced. This system consists of 3 parts: (a) a reputation system based on blockchain; (b) a two-level privacy-preserving mechanism that combines blockchain for data authentication; and (c) an intrusion detection security system. Finally, a deployment system known as Cloud-Block, Fog-Block, and Edge-Block was proposed to discuss the limitations of the present Cloud-Fog-Edge architecture. The system's performance is assessed using the BoT-IoT and TON-IoT datasets. What concerns AC, DR, PR, and F1-Score, the conclusion of the experiments ensures that the system proposed outperforms a few of the current state-of-the-art techniques. In the future, the system proposed will be extended by implementing a prototype that verifies the entire system's security parameters. This could help to enhance the overall security and privacy requirements of emerging smart cities. Finally, we will expand on this work by employing different load balancing requirements, for example, Sharding, which can improve network performance overall.

References

1. Park, E., Del Pobil, A.P., Kwon, S.J.: The role of Internet of Things (IoT) in smart cities: technology roadmap-oriented approaches. *Sustainability* **10**(5), 1388 (2018)
2. Singh, S., Sharma, P.K., Yoon, B., Shojafar, M., Cho, G.H., Ra, I.H.: Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. *Sustain. Cities Soc.* **63**, 102365 (2020)
3. Sengupta, J., Ruj, S., Bit, S.D.: A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT. *J. Netw. Comput. Appl.* **149**, 102481 (2020)

4. Jindal, A., Kumar, N., Singh, M.: A unified framework for big data acquisition, storage, and analytics for demand response management in smart cities. *Future Gener. Comput. Syst.* **108**, 921–934 (2020)
5. Kumar, P., Gupta, G.P., Tripathi, R.: TP2SF: a trustworthy privacy-preserving secured Framework for sustainable smart cities by leveraging blockchain and machine learning. *J. Syst. Arch.* **115**, 101954 (2021)
6. Čolaković, A., Hadžialić, M.: Internet of things (IoT): a review of enabling technologies, challenges, and open research issues. *Comput. Netw.* **144**, 17–39 (2018)
7. Mohanta, B.K., Jena, D., Satapathy, U., Patnaik, S.: Survey on IoT security: challenges and solution using machine learning, artificial intelligence and blockchain technology. *Internet Things* **11**, 10022 (2020)
8. Zikria, Y.B., Afzal, M.K., Kim, S.W., Marin, A., Guizani, M.: Deep learning for intelligent IoT: Opportunities, challenges and solutions (2020)
9. Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. *Science* **349**(6245), 255–260 (2015)
10. Alsheikh, M.A., Lin, S., Niyato, D., Tan, H.P.: Machine learning in wireless sensor networks: algorithms, strategies, and applications. *IEEE Commun. Surv. Tutor.* **16**(4), 1996–2018 (2014)
11. Tahsien, S.M., Karimipour, H., Spachos, P.: Machine learning based solutions for security of Internet of Things (IoT): a survey. *J. Netw. Comput. Appl.* **161**, 102630 (2020)
12. Rathore, S., Kwon, B.W., Park, J.H.: BlockSecIoTNet: blockchain-based decentralized security architecture for IoT network. *J. Netw. Comput. Appl.* **143**, 167–177 (2019)
13. Rathore, S., Pan, Y., Park, J.H.: BlockDeepNet: a blockchain-based secure deep learning for IoT network. *Sustainability* **11**(14), 3974 (2019)
14. Atlam, H.F., Walters, R.J., Wills, G.B.: Intelligence of things: opportunities & challenges. In: 3rd Cloudification of the Internet of Things (CIoT), pp. 1–6 (2018)
15. Zheng, Z., Xie, S., Dai, H.N., Chen, X., Wang, H.: Blockchain challenges and opportunities: a survey. *Int. J. Web Grid Serv.* **14**(4), 352–375 (2018)
16. Wright, K.L., Martinez, M., Chadha, U., Krishnamachari, B.: SmartEdge: a smart contract for edge computing. In: IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp. 1685–1690 (2018)
17. Swan, M.: Blockchain thinking: The brain as a DAC (decentralized autonomous organization), in Texas Bitcoin Conference, Chicago, pp. 27–29 (2015)
18. Salah, K., Rehman, M.H.U., Nizamuddin, N., Al-Fuqaha, A.: Blockchain for AI: review and open research challenges. *IEEE Access* **7**, 10127–10149 (2019)
19. Qian, Y., et al.: Towards decentralized IoT security enhancement: a blockchain approach. *Comput. Electr. Eng.* **72**, 266–273 (2018)
20. Alguliyev, R.M., Aliguliyev, R.M., Abdullayeva, F.J.: Privacy-preserving deep learning algorithm for big personal data analysis. *J. Ind. Inf. Integr.* **15**, 1–14 (2019)
21. Makhdoom, I., Zhou, I., Abolhasan, M., Lipman, J., Ni, W.: PrivySharing: a blockchain-based framework for privacy-preserving and secure data sharing in smart cities. *Comput. Secur.* **88**, 101653 (2020)
22. Koroniotis, N., Moustafa, N., Sitnikova, E., Turnbull, B.: Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset. *Future Gener. Comput. Syst.* **100**, 779–796 (2019)
23. Haider, W., Moustafa, N., Keshk, M., Fernandez, A., Choo, K.K.R., Wahab, A.: FGMCHADS: fuzzy Gaussian mixture-based coreentropy models for detecting zero-day attacks from linux systems. *Comput. Secur.* **96**, 101906 (2020)
24. Kumar, R., Tripathi, R.: Blockchain-based framework for data storage in peer-to-peer scheme using interplanetary file system. In: Handbook of Research on Blockchain Technology, pp. 35–59. Elsevier (2020)

25. Bhattacharya, S., et al.: A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU. *Electronics* **9**(2), 219 (2020)
26. Gai, K., Guo, J., Zhu, L., Yu, S.: Blockchain meets cloud computing: a survey. *IEEE Commun. Surv. Tutor.* **22**(3), 2009–2030 (2020)
27. Singh, J., Singh, J.: A survey on machine learning-based malware detection inexecutable files. *J. Syst. Arch.* **112**, 101861 (2020)
28. Gupta, R., Tanwar, S., Tyagi, S., Kumar, N.: Machine learning models for secure data analytics: a taxonomy and threat model. *Comput. Commun.* **153**, 406–440 (2020)

Image Processing and Computer Vision



Red-Channel Based Iris Segmentation for Pupil Detection

S. Bhuvaneswari^(✉)  and P. Subashini 

Department of Computer Science, Centre for Machine Learning and Intelligence (CMLI)
Avinashilingam Institute, Coimbatore, India
{bhuvaneswari_cs, subashini_cs}@avinutty.ac.in

Abstract. Iridology is the science and practice that helps to learn the health status through the examination of the structural aspects of an eye. Iris segmentation is an important phase in Iridology that aids to recognize iris signs and identify aberrant organ activity. The primary elements of iris segmentation are the detections of the iris and pupil. However, iris segmentation is greatly impacted by noises and images with low quality. The proper method for iris and pupil detection supports improving the result of iris segmentation in Iridology. This study provides a computer vision-based method to segment the human eye's iris and pupil using an alternate method. The procedure is based on the red channel information of the image with various image processing operations. Iris images from the UBIRIS.v1. Dataset, which was taken in less restrictive image acquisition conditions, is used in an experiment. This method gives a unique way for iris segmentation with an accuracy of 97.49%.

Keywords: Canny edge detection · Complementary medicine technique · Contour · Iridology · Pupil detection

1 Introduction

The human iris is as distinct as the fingerprint as there are no two irises alike. Hence, most security agencies use biotechnology or eye-based personal identification as one of their methods. The eyes can also be used as a health status indicator, according to Iridology. Reading the features of an eye to assess the health status is known as the study of Iridology (pronounced as eye-rid-ology), which is an alternative medicine technique [1]. Early diagnosis is important in the field of the health sector for successful treatments and longer survival times. Iridology serves as a pre-screening method in the medical field and helps to reveal the organs' health status by interpreting the eye's image through the advancement of technology.

The Iridologist who handles the Iridology practice on the patient uses the Iridology chart [2]. The physiology of the body is located on the iridology chart using a unique code. Many Iridology charts are introduced by many practitioners based on their own experiences and observations [3]. In these chart makers' lists, Bernard Jensen's chart is popular and internationally acceptable [4] (see Fig. 1). Computer vision is used to conclude the status of the organs by interpreting the eye based on an Iridology chart.

the feature from the iris and classify the data in addition to the K-Nearest Neighbor method. However, the cropping process was done manually. M. U. Rehman et al. [8] proposed the Ensemble learning method based on iris features. The author extracted the iris features from the Region of Interest (ROI) using Statistical Features, Gray-Level Co-Occurrence Matrix (GLCM), and Gray-Level Run Length (GLRL) Matrix to estimate the liver functionality.

Using Newton's interpolation method, F. Hernandez et al. [9] experimented to predict the early stages of Alzheimer's by examining the region of the brain represented by the structure of an iris. The Hough Transform technique was used in this investigation to find the iris and pupil, and the afflicted area was enhanced using the Logarithmic Intensity Transformation. D. C. Adelina et al. [10] suggested the GLCM for feature extraction and neural network backpropagation for classification to assess iridology and ascertain the condition of the pancreas organ. For segmenting the iris from the whole image, the author used the Hough Circle Transform method. However, the CHT failed to detect an iris portion due to excessive or inadequate light exposure. S. H. Putri et al. [11] developed a system to diagnose cholesterol by analyzing the Low-Density Lipoprotein (LDL) level. The researcher used Image processing techniques for preprocessing and a Convolution neural network for classification. To localize the iris province on the image, the Circle Hough Transform is used. However, there is a need for some restrictions to isolate the iris due to the greyish-white colored ring (Cholesterol ring) which is surrounded by the iris region. Onal M.N. et al. [12] proposed a hybrid method which is incorporating image processing and deep learning to diagnose diabetes based on iris images. The author used Daugman's Integro-Differential Operator (IDO) technique to distinguish the iris from the rest of the image and to locate the pancreatic region using the Iridology chart [13]. After segmenting, the entire iris area, the cropped ROI, is categorized by various deep-learning networks, and performances are assessed separately.

The primary goal of all previous techniques was to identify specific body organ malfunctions. In this study, the results of an iris detection are evaluated using a standard grayscale image and a different approach that uses red-channel information of an image.

3 Experimental Methodology

This study tries to provide an alternate method for segmenting the iris. Lately, most experiments have been made using Near Infrared Images (NIR) as input. Because the pupil is not stimulated by NIR illumination. Therefore, during image acquisition, pupil dynamics do not significantly affect the texture of the iris. Moreover, in contrast to the color irises, the texture of dark-colored irises is more discernible in the Infrared spectrum [14]. NIR camera help to obtain a clear image of the iris, thus the pupil and iris can be seen with great contrast in NIR images [15, 16]. As a result, pupil detection in NIR images is straightforward. However, when compared to NIR images, color images are noticeably noisier. [17]. Thus, segmenting the iris in color images is not an easy task.

The human iris has various iris colors, patterns, and pigments. In addition, the brightness of the light will dilate or shrink pupil size [18]. Thus, iris recognition will be impacted by differences in iris structure, iris color, and pupil dilation or contraction. The study represents a single-color spectrum approach based on the red-channel information

of an image during iris localization to improve its efficiency. Figure 2 depicts a red channel-based iris detection workflow.

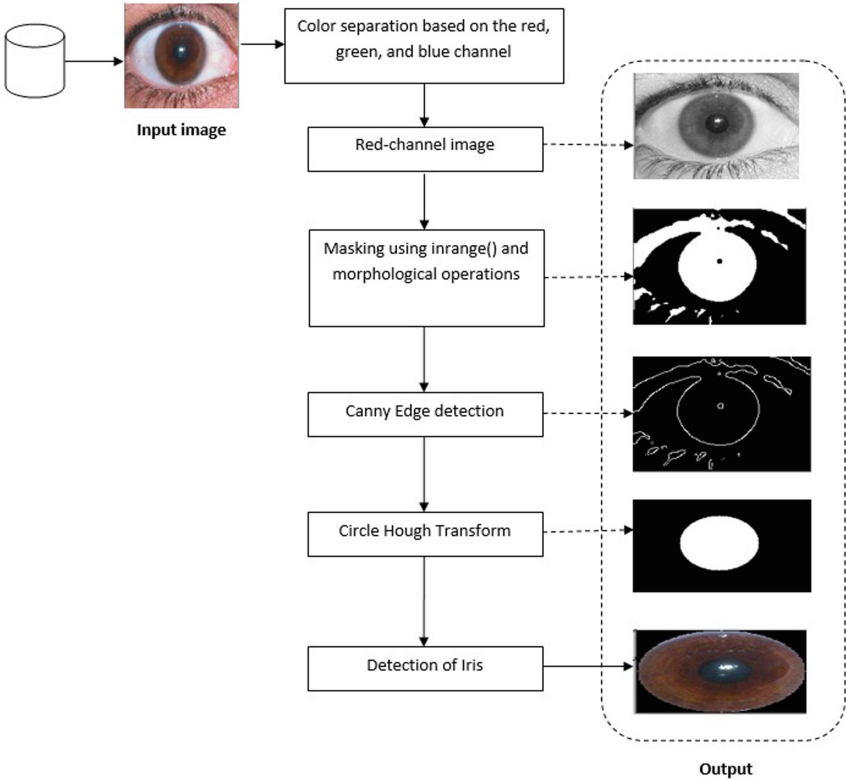


Fig. 2. Workflow for iris recognition using red-channel image

3.1 Iris Localization

In the human eye, the common feature is pupil color. The pupil color is always dark in the eye, unlike the iris. Hence, the authors initially try to detect the pupil from the human eye, though the illumination or an uncontrolled environment will affect the pupil’s darkness. It will make difficulties in pupil detection. Hence, this study aims to locate the iris of an eye before detecting the pupil region.

3.1.1 Pre-processing

Color Conversion

With data collected from the UBIRIS.v1 database, the method was evaluated. The dataset contained RGB-formatted images that could be accessed [19]. So, the Color

conversion will help to read more details about the input data. The human iris is differentiated by its various color. It needs some common color conversion method that is suitable for research work. Grayscale image conversion is the most popular technique due to its range of color, and low computation power. Apart from the NIR input image, in iris recognition most of the authors prefer the grayscale image for iris detection and pupil detection, to make the complex task simple. The Grayscale image range comes under three categories; one is black, the second is white, final is a gradient between black and white. Those ranges will help to detect or separate the required region of the input image.

In the RGB image, each pixel represents three channels. The color of the pixel is decided by the combination of red, green, and blue colors. An area that contributes more red color in the original image produces more brightness in the red channel image, if not look darker. This applies to the remaining two channels. So, the separated color channel images will look like the grayscale image, for the reason the pixel of these images has only one channel like the grayscale image. This study used red channel-based color images to evaluate and compare the outcomes of iris segmentation with conventional grayscale images.

Noise Removal

The acquired input data may be affected by the noise. The noise filtering techniques help to reduce the noise, and the filtered results aid to increase accuracy (see Fig. 3). The median filter aids to remove the impulse noise from an input image. Therefore, a median filter is a feasible option for reducing noise from the input image of the human eye without degrading the sharpness of the image [20].



Fig. 3. a) Input image b) Before noise reduction c) After noise reduction.

3.1.2 Edge Detection

In the eye, the pupil represents the prominent level of black color, and the sclera represents the major level of white or bright color. The colored area of the eye is termed as iris. These distinctions aid in separating an iris from the eye. The inRange function of thresholding is used to increase the visibility of the iris part. The inRange function detects the region, based on the pixel values. It returns the number of elements that lies between the given upper and lower bounds.

The inRange emphasizes the iris portion. To improve this result, the erosion operation is applied to the result of inRange method. One of the techniques used in morphological operations is erosion. It will compare the kernel against the matrix of the input image. If there is a perfect match, it assigns the binary value 1 (bright), else assigns the value

0(black). As a result, the increased black will make the necessary part of an image more visible and reduce reflection. The final image makes it simpler to see the edges. For better edge identification erosion operations are used in conjunction with thresholding operations using the `inRange`. The edges in the resulting binary image are identified using the Canny edge detection method. To identify the strong edges in the input image, the Canny edge detection technique performs the five steps as follows: 1) With the use of a Gaussian filter, the Canny edge detection technique smooths and eliminates distortion from the input image to generate strong edge pixels. 2) Identifies the horizontal, vertical, and diagonal edges in the blurred image by detecting the intensity gradient of the image. 3) Then it applies the edge thinning technique to find the sharpest changes in intensity value locations. 4) After applying the non-maximum suppression even though the resultant image may contain thin edge pixels due to the noise and color variation. To suppress this weak edge, the Canny algorithm applies the double threshold. 5) Finally strong edge pixels by using hysteresis are traced.

To do noise reduction, the Canny edge algorithm uses a Gaussian kernel (3x3, 5x5, 7x7, 9x9, etc.) to perform the Convolution technique on the image. The kernel size depends on the required blur range.

A Gaussian filter kernel size's equation is,

$$G_{\sigma} = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (1)$$

The following equation is used by the Canny edge algorithm to calculate the gradient value after the image has been smoothed:

$$|G| = \sqrt{G_x^2 + G_y^2} \quad (2)$$

$$\theta(x, y) = \arctan\left(\frac{G_y}{G_x}\right) \quad (3)$$

3.1.3 Circle Hough Transform (CHT)

Different kinds of object detection and recognition algorithms are dealt with in image processing. The iris looks in the shape of an ellipse. Additionally, CHT is the widely used computer vision algorithm for locating the circle's center coordinate values [21, 22]. Thus, CHT is the desired method to detect an iris.

To detect the circle, three parameters (a, b, R) are required mathematically. The circle's center points are a and b, and its radius is r. It will produce the edges $[(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)]$ using these three parameters around the given center point. The following three parameters can be used to define a circle:

$$X = a + R\cos(\theta) \quad (4)$$

$$Y = b + R\sin(\theta) \quad (5)$$

This is the principle upon which the Circle Hough Transform (CHT) operates. When a circle's center point is uncertain, the radius is taken into account and generates a circle

on the edge which is given by the Canny edge detection algorithm. The CHT algorithm moves along the strong edge (x, y) and plots the circles, then compute the coordinates of the center of all the circles. It considers the high voting feature point from the accumulator array as the center point of the circle (see Fig. 4).

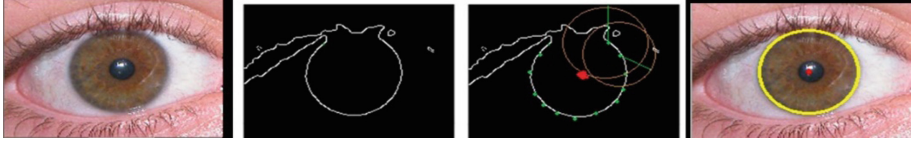


Fig. 4. Iris identification. a) Input image b) Edges of an image c) Plotting circles along edges with predefined radius d) Outcome with a high voted center point

3.2 Pupil Detection

Iris detection makes pupil detection simpler. When the iris pigments are dispersed across the pupil boundary in grayscale images, the threshold value occasionally impacts the iris area (see Fig. 5). The pupil was therefore identified in this analysis using RGB images. In comparison to a grayscale image, it produces better outcomes.

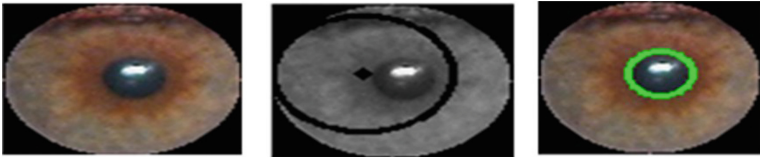


Fig. 5. Pupil detection. a) Input image b) Grayscale image c) RGB image

In this method, the binary representation of an image is changed by selecting and masking the pupil ranges. Then the Morphological closing operation is applied to the masked regions. The reflection noise in the iris images can be removed through morphological processing [23]. The Morphological closing operation helps to smooth boundaries and fill small holes or small black points on objects. Later, the contour method is applied to the result of the Morphological operation. Then the edge detection using Canny is then applied to the generated image.

The Canny edge detector finds the sudden intensity change's edges. It helps to improve the outcome of the Circle Hough Transform. It can be difficult to determine the best value for the lower and upper thresholds in the Canny edge detection algorithm when images are taken in various lighting situations. The auto canny detecting technique is applied to make it effective. The mean value of the binary images that have been masked is used along with the sigma value that works best for all the images. The manual tweaking that depends on the images is avoided by the auto Canny detector. In order to find the pupil, the Circle Hough Transform is applied to the images after creating the edges.

4 Results and Analysis

4.1 Dataset

Images from the UBIRIS.v1 dataset were used as input to assess its performance. The UBIRIS.v1 database encompasses a collection of 1877 images that were captured from 241 various populations using a Nikon E5700 camera. The major purpose of this approach is to detect the pupil through iris segmentation. This study analyzes the two distinct types of input images: 1) Grayscale image and 2) Single color spectrum based on red-channel information of an image which is suggested by this paper.

4.2 Iris Localization

The results are good for both the standard and red channel images. However, the red channel-based images still provide the optimum solution. To evaluate the effectiveness of iris segmentation in both conventional grayscale and red channel images, the dataset's input images are chosen at random (see Fig. 6(a)). The results of an iris localization analysis for the red-channel image and the typical grayscale image are exhibited in Figs. 6(b) and 6(c), correspondingly.

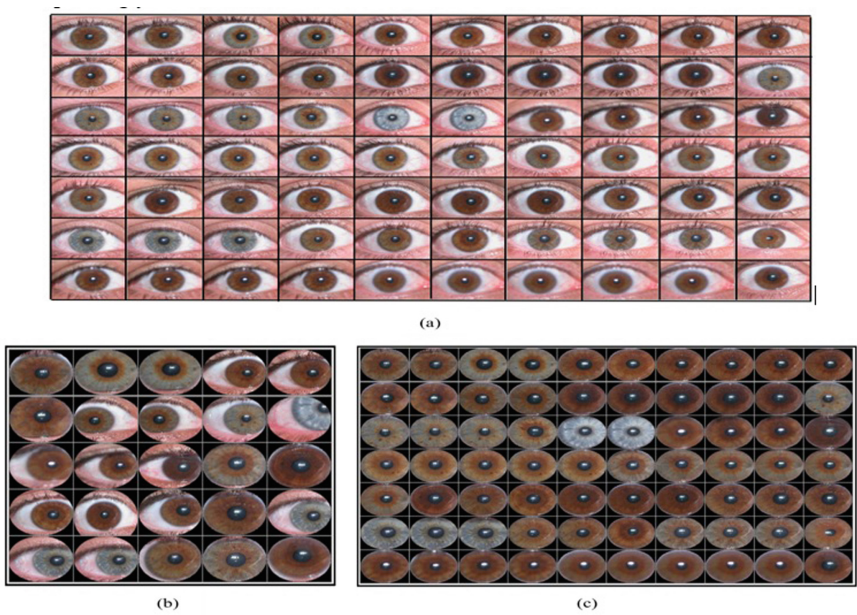


Fig. 6. Iris detection. a) Input image b) Grayscale image c) Red-channel image

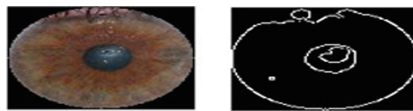
The results of the iris's segmentation accuracy using grayscale and red-channel images are shown in Table 1. The images with the red channel exhibit the highest accuracy when compared to grayscale images. Through this experiment, it was determined that red channel information may more accurately segment the iris than grayscale images.

Table 1. Accuracy of different methods in detection on the same dataset

Total images	Red-channel image		Grayscale image	
	Detected	Poor Detection	Detected	Poor Detection
205 (Controlled Environment images)	202	3	140	65
34 (Less controlled Environment images)	31	3	27	7
239	233	6	167	72
Accuracy	97.49%		69.87%	

4.3 Pupil Detection

The pupil detection is achieved using RGB images. The pupil area is masked using the minimum and greatest range of pupil intensity values. However, the reflection makes pupil detection difficult. In order to identify the reflection area, the threshold operation is used. It is found that the reflection range exceeds 200 after considering the various input images. The inpaint approach is used to provide a solution to fill the reflection area after the reflection portion has been identified. However, it creates a watermark in the painted area. These marks will make edge detection more challenging (see Fig. 7). Therefore, to solve this issue, Morphological operations are used. As a result, the experiment uses the Morphological closing operation along with the filling contour method to reduce reflection.

**Fig. 7.** Marks after reflection removal

In terms of object detection and segmentation, the contour performs well. It defines the area of the image that is within the specified intensity range. This concept is utilized in the filling contour approach for reflection removal. Thus, the contour area is detected and filled or painted by the bright intensity value. The reflection area and the pupil area can be integrated, according to this conception. It will help to improve the pupil region's detection (see Fig. 8).

The results of the segmented iris images are randomly chosen to test the effectiveness of pupil detection using the filling contour approach. The outcome of pupil detection is displayed in Fig. 9. In contrast to a technique without a contour method, the filling contour method offers the best result.

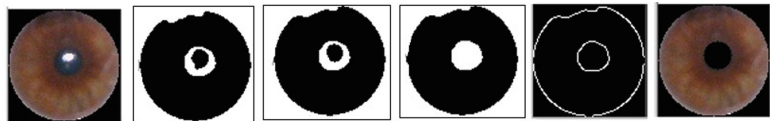


Fig. 8. Pupil detection using the filling contour method

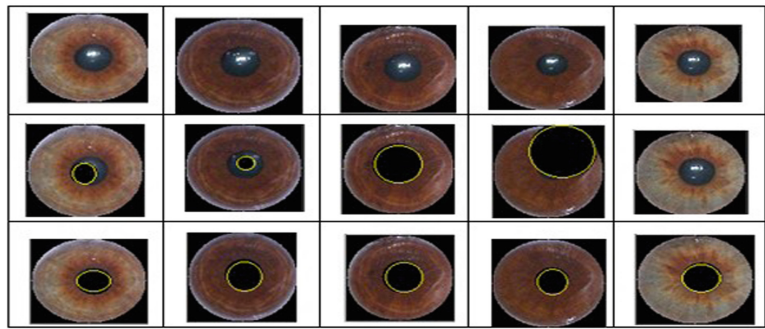


Fig. 9. Pupil detection. Top row: Segmented Iris images. Middle row: Output images obtained by the non-contour method. Bottom row: Output images obtained by the filling contour method

5 Conclusion

This study provides an alternative approach to segmenting the iris and pupil in Iridology. The input image of the eye is subjected to various types of image processing algorithms based on red channel information in order to identify the pupil. The filling contour method is applied to minimize the reflection during pupil detection. This study experimented with color images, which are useful for analyzing visible wavelength images. Thus, this distinctive and useful method will be advantageous for color images taken with smartphones. This work also provides a useful analysis of the grayscale images and red channel images as part of the iris segmentation procedure. In this paper’s experimental study, only publicly accessible iris images were used. Future research will therefore concentrate on real-time images with improved versions of various types of algorithms and classification techniques in an effort to further reduce mistake rates.

Acknowledgment. This work has been supported by the Centre for Machine Learning and Intelligence (CMLI) funded by the Department of Science and Technology (DST-CURIE).

References

1. Jensen, B.: Visions of Health: Understanding Iridology. Penguin, New York (1991)
2. Iridology Chart: <https://myinfiniteiris.com/product/updated-wallet-size-iridology-chart/>
3. Pau, G.: The Foundations of Iridology: The Eyes as the Key to Your Genetic Health Profile. Simon and Schuster, New York (2019)

4. Esteves, R.B., Morero, J.A., Pereira, S.D., Mendes, K.D., Hegadoren, K.M., Cardoso, L.: Parameters to increase the quality of iridology studies: a scoping review. *Eur. J. Integr. Med.* **43**, 101311 (2021)
5. Huda, A.L.: Iris detection using morphology. *J. Univ. Babylon* **22**(9), 2277–2282 (2014)
6. Permatasari, L.I., Novianty, A., Purboyo, T.W.: Heart disorder detection based on computerized iridology using support vector machine. In: 2016 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC), pp. 157–161 (2016)
7. Lestari, R.F., Nugroho, H.A., Ardiyanto, I.: Liver detection based on iridology using local binary pattern extraction. In: 2019 2nd International Conference on Bioinformatics, Biotechnology and Biomedical Engineering (BioMIC) - Bioinformatics and Biomedical Engineering. no. 1, pp. 1–6 (2019)
8. Rehman, M.U., Najam, S., Khalid, S., et al.: Infrared sensing based non-invasive initial diagnosis of chronic liver disease using ensemble learning. *IEEE Sens. J.* **21**, 19395–19406 (2021)
9. Hernandez, F., Vega, R., Tapia, F., Morocho, D., Fuertes, W.: Early detection of Alzheimer's using digital image processing through iridology, an alternative method. In: 2018 13th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–7. IEEE (2018)
10. Adelina, D.C., Sigit, R., Harsono, T., Rochmad, M.: Identification of diabetes in pancreatic organs using iridology. In: 2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), pp. 114–119 (2017)
11. Putri, S.H., Saputro, A.H.: Design of convolutional neural network modeling for low-density lipoprotein (LDL) levels measurement based on iridology. In: 2020 4th International Conference on Informatics and Computational Sciences (ICICoS), pp. 1–5 (2020)
12. Onal, M.N., Guraksin, G.E., Duman, R.: Convolutional neural network-based diabetes diagnostic system via iridology technique. *Multimedia Tools Appl.* **82**(1), 173–194 (2023)
13. Daugman, J.G.: How iris recognition works. *IEEE Trans. Circuits Syst. Video Technol.* **14**, 21–30 (2004)
14. Bobeldyk, D., Ross, A.A.: Predicting eye color from near infrared iris images. In: 2018 International Conference on Biometrics (ICB), pp. 104–110 (2018)
15. Lee, M.B., Kang, J.K., Yoon, H.S., Park, K.R.: Enhanced Iris recognition method by generative adversarial network-based image reconstruction. *IEEE Access* **9**, 10120–10135 (2021)
16. Wang, Q., Zhipeng, L., Tong, S., Yang, Y., Zhang, X.: Efficient iris localization via optimization model. *Math. Probl. Eng.* **2017**, 1–9 (2017)
17. Hu, Y., Sirlantzis, K., Howells, G.: Improving colour iris segmentation using a model selection technique. *Pattern Recognit. Lett.* **57**, 24–32 (2015)
18. Mathot, S.: Pupillometry: psychology, physiology, and function. *J. Cogn.* **1**(1), 16 (2018)
19. Proença, H., Alexandre, L.A.: UBIRIS: a noisy iris image database. In: Roli, F., Vitulano, S. (eds.) *Image Analysis and Processing – ICIAP 2005. Lecture Notes in Computer Science*, vol. 3617, pp. 970–977. Springer, Heidelberg (2005). https://doi.org/10.1007/11553595_119
20. Gangwar, A.K., Joshi, A., Singh, A., Alonso-Fernandez, F., Bigun, J.: IrisSeg: a fast and robust iris segmentation framework for non-ideal iris images. In: 2016 International Conference on Biometrics (ICB), pp. 1–8 (2016)
21. Samant, P., Agarwal, R.: Comparative analysis of classification-based algorithms for diabetes diagnosis using iris images. *J. Med. Eng. Technol.* **42**, 35–42 (2018)
22. Jesus, R.J., Maximo, L., Pinto Elías, R., Gabriel, G.: Methodology for Iris scanning through smartphones. In: 2016 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 861–864 (2016)
23. Malgheet, J.R., Manshor, N.B., Affendey, L.S.: Iris recognition development techniques: a comprehensive review. *Complexity* **2021**, 1–32 (2021)



Low Overhead Routing in a Lightweight Routing Protocol

Maryem Lachgar^(✉), Ibtissam Larhlimi, Anouar Darif, Hicham Ouchitachen,
and Hicham Mouncif

Department of Mathematics and Informatic, Polydisciplinary Faculty,
University Sultan Moulay Slimane, Beni-Mellal, Morocco
maryamlachgar96@gmail.com, ibtissam.larhlimifbp@usms.ac.ma,
h.mouncif@usms.ma

Abstract. Reducing the amount of data transmitted over the network will reduce energy consumption. One of the key issues facing routing protocols in wireless sensor networks is link breakage. Sometimes it causes the traffic to halt entirely or be temporarily suspended. Due to this, the routing protocol must be designed to allow for link breaks in order to identify an other route or manage the issue locally by offering a local link repair that replaces only the damaged portion of the route with the least amount of routing overhead possible to improve energy efficiency. This work introduces the local repair with various ways to lower the routing overhead in the Lightweight Routing Protocol (LRP).

Keywords: Wireless sensor network · Routing · Local repair · LRP

1 Introduction

Routing is important for discovering, establishing, and maintaining routes along which data packets can be routed in multi-hop wireless sensor networks [1]. This necessity results from the fact that networks are typically too large for the human mind to fully comprehend, that wireless link quality is highly variable, that nodes may go offline, that new nodes may join the network, and that nodes may move around, and that calculations to find the best route among the options are a meticulous task. Therefore, the routing protocol must ensure that, in the event of a topology change, the routing structure will be properly adjusted without introducing loops. Additionally, it must add little overhead and consider the limitations of the energy resources.

LRP, or Lightweight Routing Protocol, combines RPL's collection tree or DODAG topology to extract traffic from the network via the sink with LOADng's reactive route construction to distribute traffic within the network. In order to generate as little traffic as possible, this combination aims to be as simple as possible.

This paper start with related works (Sect. 2). (Section 3) introduces the wireless sensor network, after that we go to a description of the protocol LRP in

(Sect. 4). Then, in (Sect. 5), we define the various strategies for reducing routing overhead. Results are presented in (Sect. 6), along with discussions of how each scenario performed, and a conclusion is provided (Sect. 7).

2 Related Works

[2] Take into account the essential issue of routing in wireless sensor networks. In order to conserve energy, routing protocols must handle changes in topology while reducing routing overhead, particularly the number of broadcasts.

In traditional works, several forms of clustering and data aggregation algorithms are created, with the goal of enhancing both energy conservation and network lifetime. However, it faces increased processing complexity, inefficient data routing, excessive regulating overhead, and decreased reliability. This work [3] investigates the development of a novel energy efficient mechanism by combining the features of advanced clustering, path selection, and data aggregation approaches.

Time-To-Live based Expanding Ring Search is one of the concepts that is frequently used to reduce broadcast overhead and flooding strategies for routing in wireless networks. In this research [4], an improved routing algorithm called Energy Efficient ERS (E2-ERS) is proposed. E2-ERS implements an effective route discovery strategy by lowering the total amount of transmissions of RREQ during route discovery for multicast in MANET. Additionally, it contrasts the original ERS technique with the suggested Energy Efficient ERS.

3 Wireless Sensor Network

Application-oriented wireless sensor networks (WSNs) are primarily employed in the following industries: border security, Water Monitoring, disaster management, and Military Applications [5]. The widespread use of these sensors has led to the development of several new services and applications that have the potential to improve our quality of life. On the other hand, WSNs have limited resources in terms of processing capabilities, memory, and power, whereas the

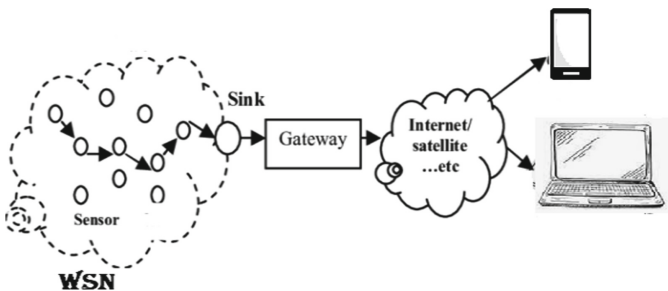


Fig. 1. A wireless sensor network's architecture.

sink is a resource-rich device with a constant power supply and communication capabilities. As a result, the sink can communicate with other computers over different networks such as a local area network, wireless local area network, wireless personal area network, and the Internet [6]. Figure 1 shows the architecture of a typical WSN.

Because of their distinct properties, typical network solutions and protocols cannot be applied to this type of environment. In recent years, standardization efforts have resulted in the development of various protocols adapted to the confined environment of the motes employed, such as the LRP routing protocol, which is crucial for enhancing performance in a variety of sensor-based applications. In the following sections, we will look at the method employed by LRP to reduce control message overflow.

4 LRP - Protocol Overview

4.1 Tree of Collections

LRP arranges the nodes into a collection tree in order to make it simple to extract data produced by nodes outside of the sensor network. This sort of traffic employs a single route known as the default route, which is installed in the routing tables of each individual node and forms a tree with a root on the sink, and is routed in the same manner regardless of the source of the packet. The distributed Bellman-Ford algorithm is used to calculate the positions in the collection tree [7].

The DIO propagate over the network after the sink starts to build the tree. When a node receives a DIO, it compares its present position to the one it would have if it had chosen the message's sender as its successor. It has located a better spot in the tree if the suggested path enables it to approach the sink.

Figure 2 illustrates the Bellman-Ford algorithm's steps for creating the collection tree.

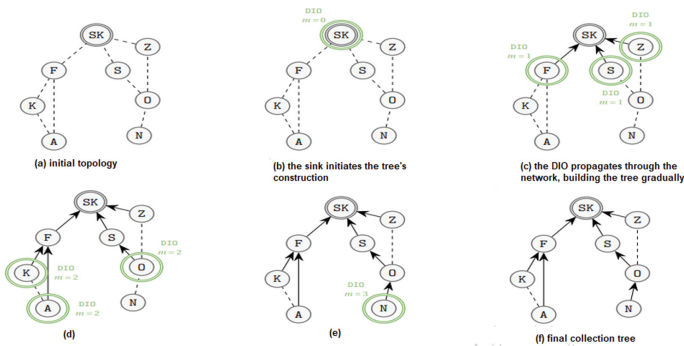


Fig. 2. Step-by-step distributed Bellman-Ford method for collection tree construction.

4.2 Host Routes

LRP enables the nodes to create host routes as well, which distribute traffic around the network. A proactive approach is taken in the initial building of the host routes, while the search for a host inside the network is done reactively.

To prevent loops from developing, the LRP's implicit data path validation algorithm detects routing loops as soon as they are used [2].

4.3 Local Repair

LRP requires that a node shouldn't ever take a route that would move it further away from the sink than its existing location in order to ensure that there are no loops into the network. There are, however, some special cases that are troublesome, such as when a node looks to identify a new successor among its neighbors but is unable to connect to the sink at all because none of its neighbors are close enough to the sink to provide a suitable path. The tree needs to be repaired at this time.

Only a few nodes are affected by local repair when it is carried out locally. It is activated When a node loses all of its successors and must use a node in its own sub-tree as success [8]. Any node in the network has the ability to trigger it when necessary. This is performed in two steps: looking for a path from the detached node to the sink with a BRK (Break) message and reconstructing a portion of the tree from the sink without requiring a routing loop with a UPD (Update) message.

- Find an alternative route to the sink: The BRK (BRAEK) message is sent up the network using the default routes as a unicast along the collection tree outside and a multicast inside the node's subtree. This phase serves two purposes: informing the sink that a local repair is necessary and establishing a connection between the sink and the disconnected node (see Fig. 3)
- the collection tree updated without introducing a loop The sink sends a UPD (Update) message in response to receiving the BRK message. The nodes in the first step store the path that the Break msg took to get to the sink as it

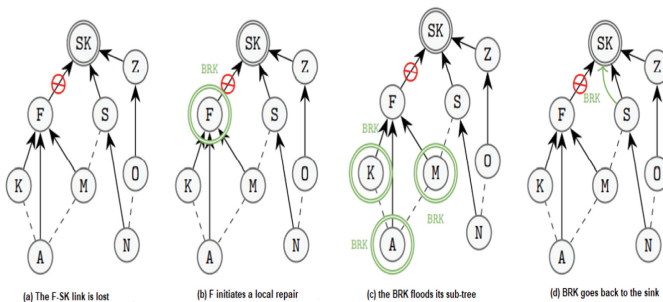


Fig. 3. Local repair “find another way to the sink”.

passed it. The UPD message travels along the same path but in the reverse direction from node to node. A node switches to utilizing the UPD sender as successor when it receives a UPD message. So, the path is reconstructed, and the disconnected node locates a new successor (Fig. 4).

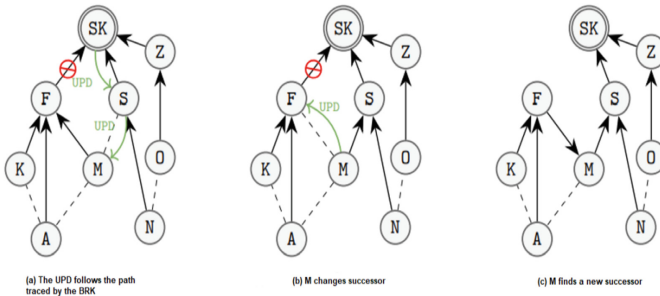


Fig. 4. Local repair “Updating the collection tree”.

5 Reduced Routing Overhead

5.1 Reduce BRK Flooding

A BRK message is distributed throughout the detached sub-tree when a detached node initiates a local repair. In many situations, the repair routes could be located through a not too distant predecessor, therefore flooding the entire sub-tree is typically not needed.

LRP suggests using an expanding ring search scheme (ERS) [9] to reduce the total number of messages broadcast. The maximum distance that the BRK is permitted to travel in the sub-tree is specified in the ring-size field of the BRK message. At zero, it prevents the BRK packet from being retransmitted. It is decremented at each hop. Nothing prevents the detached node from increasing the ring size if the collection tree cannot be repaired using the closest nodes.

The BRK range restriction only applies within the subtree. When the BRK is outside, the path to the sink has been discovered, and the message has become unicast. After then, the ring-size field is ignored.

5.2 Spontaneous Response to a DIO

To understand this mechanism we take two neighboring nodes A and B.

When B at a distance n from the sink, broadcasts a DIO describing a distance to the sink that is greater than A. The latter spontaneously responds by sending back a unicast DIO message detailing its own distance, prompting neighbor B to change its successor.

With this mechanism, B may easily explore its neighborhood and find potential successors by publishing its own distance to the sink in an DIO message. It will only get responses from neighbors who can be used as successors.

5.3 Host Routes Repair

Several nodes in a local repair change successor. The following issue is the collection of nodes in its subtree, whose host routes are generally no longer usable.

Once detecting that host routes are broken, the sink will hunt for their destinations reactively by flooding numerous RREQ message across the entire network. This could result in a substantial amount of broadcast traffic, which we wish to prevent. Proactively constructing the host route by sending an RREP message on behalf of the nodes in its subtree is a primitive method because it cannot change the sequence number it received with the RREP message. As a result, it is critical that the host generate a new sequence number, leaving the previous one obsolete. This is the aim of the constrained RREQ, as seen in the image below.

When a node receives restricted RREQ, it sends it to its successor; however, the CONFINED flag specifies that it should only be processed when a node receives it from its successor. F broadcasts the circumstance to its full sub-tree via confined RREQ, then Spontaneous RREP travel up the collection tree, and finally All host routes are rebuilt (Figs. 5 and 6).

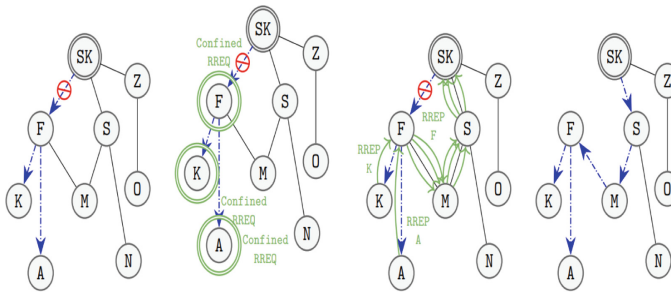


Fig. 5. Host routes repair.

6 Results and Simulation

This experiment’s goal is to assess the effectiveness of the mechanisms described in Sects. 5.1, 5.2, and 5.3 by contrasting the two implementations of the LRP routing protocol “o-LRP and LRP” in a real situation. In this case, several functionalities in o-LRP are deactivated (Table 1).

Table 1. Node parameters

Parameter	Value
Random delay for packet transmission following a broadcast	0.5 s
Frequency of neighborhood probing when disconnected	5 min
Blacklist timeout	10 min
Metric type	Hop count

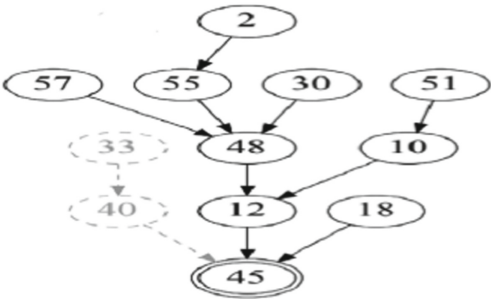


Fig. 6. Collection tree of the experimentation.

Table 2. The total number of messages sent throughout the experiment

		o-LRP	LRP
Broadcast	DIO	37	4
	BRK	7	5
	RREQ	11	3
Unicast	DIO	n/a	1
	BRK	20	4
	UPD	12	5
	HELLO	n/a	2
	RERR	3	–
	RREP	25	17

O-LRP Simulation

Node 40 uses the local repair procedure because it is unable to access the sink, which causes DIO, BRK, and UPD messages to first arrive shortly after 15 min. The results of this experiment are not very promising. A network saturation issue does in fact occur because another node failed to proactively create its host route (Fig. 7).

In o-LRP (the original version) a flood of RREQ and RREP messages is caused by the fact that the detached node must be eliminated and rebuilt reactively.

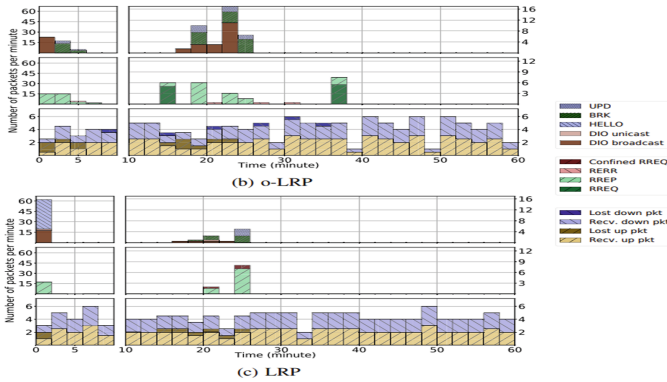


Fig. 7. 11 client nodes and 1 sink in 60-min trials.

Since in o-LRP (the original version) the detached node must be removed and rebuilt reactively, this results in a flood of RREQ and RREP messages.

LRP Simulation

Although host routes that contact the repaired subtree are not necessarily updated at this point, local repair initially functions similarly to o-LRP. As a result, it uses the confined RREQ method to proactively repair all host routes.

Both studies effectively reestablish connectivity in a matter of minutes. However, as seen in Table 2, o-LRP sends more RREQ, RREP, and RERR messages than LRP, resulting in more control traffic and energy consumption.

7 Conclusion

With the application of the suggested techniques, LRP may be regarded as one of the lightest and most effective routing protocols. Few control messages are sent by nodes. LRP often uses a proactive initialization of the collection tree followed by a cessation of message broadcasting. When links break, LRP immediately responds to rebuild the tree and proactively update these host routes to prevent reactive rebuilding and a resulting network flood. Links failure is detected during data packet transmission.

The Perspectives

The objective function specifies how a node chooses the best path given routing metrics and constraints. It is regarded as the most important aspect in determining a node's rank and also defines how nodes choose parents.

The next step is to look for and propose an objective function that will allow LRP to provide more reliable findings in order to estimate precisely the optimal path when only a few exchanges have occurred on it.

References

1. H., El-Sayed, H., Al Bayatti, H.: Comparisons of some multi-hop routing protocols in wireless sensor networks. *Inf. Sci. Lett.* **10** (2021)
2. H., J., Audeoud, M., Krol, M., Heusse, A., Duda: Low overhead loop-free routing in wireless sensor networks. In: 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 443–451 (2015). <https://doi.org/10.1109/WiMOB.2015.7347996>
3. Arunachalam, G.S., Vimal, S., Ramalingam, R.G.: Nanjappan: a classy energy efficient spider monkey optimization based clustering and data aggregation models for wireless sensor network. *Concurrency and Computation: Practice and Experience* **35** (2023). <https://doi.org/10.1002/cpe.7492>
4. Kinjal, U.: Adhvaryu: energy efficient ers for multicast routing in manet. International Conference on Recent Advances in Energy-Efficient Computing and Communication (ICRAECC) (2019). <https://doi.org/10.1109/ICRAECC43874.2019.8995085>
5. G., Preeti, S., Tripathi, S., Singh: Energy efficient hotspot problem mitigation techniques using multiple mobile sink in heterogeneous wireless sensor network. *Int. J. Commun. Syst.* **33**(18) (2020)
6. Mehto, A., Tapaswi, S., Pattanaik, K.K.: A review on rendezvous based data acquisition methods in wireless sensor networks with mobile sink. *Wirel. Networks*
7. Bellman, R.: On a routing problem. *Quarterly Appl. Math.* **16**(1), 87–90 (1958)
8. Henry-Joseph, A., Martin, Heusse, H.: Experimental comparison of routing protocols for wireless sensors networks: routing overhead and asymmetric links. In: 29th International Teletraffic Congress (2017)
9. Duy, N., et al.: An expending ring search algorithm for mobile adhoc networks. In: The 2010 International Conference on Advanced Technologies for Communications, pp. 39–44 (2010). <https://doi.org/10.1109/ATC.2010.5672725>



Identification and Localization COVID-19 Abnormalities on Chest Radiographs

Van Tien Pham^(✉) and Thanh Phuong Nguyen

Université de Toulon, Aix Marseille Univ, CNRS, LIS, Marseille, France
van-tien-pham@etud.univ-tln.fr, tpnguyen@univ-tln.fr

Abstract. Solutions to screen and diagnose positive patients for the SARS-CoV-2 promptly and efficiently are critical in the context of the COVID-19 pandemic's complex evolution. Recent researches have demonstrated the efficiency of deep learning and particularly convolutional neural networks (CNNs) in classifying and detecting lung disease-related lesions from radiographs. This paper presents a solution using ensemble learning techniques on advanced CNNs to classify as well as localize COVID-19-related abnormalities in radiographs. Two classifiers including EfficientNetV2 and NFNet are combined with three detectors, DETR, YOLOv7 and EfficientDet. Along with gathering and training the model on a large number of datasets, image augmentation and cross validation are also addressed. Since then, this study has shown promising results and has received excellent marks in the Society for Imaging Informatics in Medicine's competition. The analysis in model selection for the trade-off between speed and accuracy is also given.

Keywords: Chest X-ray · COVID-19 radiographs · Classification and detection · Deep neural network · Medical imaging

1 Introduction

As the global coronavirus disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), spreads throughout the world, radiology is becoming increasingly important in giving clinical insights to help in illness diagnosis, treatment, and management. An increasing interest and use of chest X-ray (CXR) imaging has been demonstrated in a number of recent research [6], with some studies predicting a higher dependence on portable CXR and the high value of portable CXR for critically sick patients. CXR imaging systems are more widely available across the world than computed tomography (CT) scanners due to their lower cost and shorter decontamination periods. For identifying SARS-CoV-2, CXR imaging can be beneficial for patients with negative reverse transcription-polymerase chain reaction (RT-PCR) findings. As a result of the popularity of chest radiology imaging systems in current healthcare systems and the availability of portable equipment, radiography examinations may be performed more quickly and with more availability, making this a good

T. P. Nguyen—Supported by project ANR ASTRID ROV-Chasseur.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
A. E. Hassanien et al. (Eds.): AICV 2023, LNDECT 164, pp. 251–261, 2023.
https://doi.org/10.1007/978-3-031-27762-7_24

complement to RT-PCR testing [14]. Hence, CXR is an essential component of certain screening methods that have been presented. However, one of the most significant obstacles is the requirement for trained radiologists to interpret radiography imaging, as the visual indications might be subtle. While standards exist to assist radiologists in distinguishing COVID-19 from other forms of infection, radiologists' judgments vary. As a result, computer-aided diagnostic tools that can help radiologists analyze radiography pictures more quickly and reliably to discover COVID-19 patients are greatly sought.

Motivated by the urgent need of solutions to fight against the COVID-19 pandemic, this work tackles the problem of classification and detection COVID-19 abnormalities on chest radiographs via ensemble of deep CNNs. In particular, the radiographs will be categorized as 4 types: negative for pneumonia or typical, indeterminate, or atypical for COVID-19. Furthermore, the appearance of glass ground opacity and the extent to which they cover lung areas in the chest X-ray were also detected to help radiologists assess the COVID-19 pneumonia. Models from this research are evaluated in the competition [10] hosted by the Society for Imaging Informatics in Medicine, the Foundation for the Promotion of Health and Biomedical Research of Valencia Region and the Radiological Society of North America (SIIM-FISABIO-RSNA) on Kaggle. Our main contributions are:

- First, an in-depth exploratory data analysis is conducted on the SIIM-FISABIO-RSNA COVID-19 dataset [10] to reflect its properties and then suggest appropriate methods. Also, many open access benchmark datasets are collected and reviewed in order to pre-train models.
- Second, this work proposes a novel framework covering well-known object classifiers and detectors. Many methods, including transfer learning, data augmentation, and ensemble learning, are used to enhance the performance of models. This study also suggests using a classifier as a initial filter to eliminate false positive samples before passing to detectors. Separating the pulmonary approach proven to be an effective method concurrently.
- Third, several comparative evaluations are analyzed in terms of accuracy and computing cost. The recommendations for CXR-diagnosis practical applications are then highlighted.

2 Related Works

Aim to interpret radiographic images better, many machine learning systems have been proposed [6], and the results show that the accuracy of radiographic imaging in detecting patients infected with COVID-19 is very promising while the cost of time is much reduced. Wang et al. [22] customized deep CNN models for identifying normal, pneumonia, and COVID-19 patients in the COVID-Net project. To provide systems aiding COVID-19 diagnosis, a variety of models [5] have been trained using X-ray and CT images. Shi et al. [15] retrieved a list of characteristics from CT images, including volume, histogram, and surface features. The open source COVID-19 Image Data Collection [4] is an example of

current initiatives to advocate for open access and solutions for radiography-driven COVID-19 case detection. Bilateral abnormalities, ground glass opacity, and interstitial abnormalities have all been identified in CXR radiograph of SARS-CoV-2 positive patients [7]. Similarly, in this study, we follow these concepts but also the rule of the SIIM-FISABIO-RSNA’s competition with the 2 main tasks. For each test study in classification task, it’s required to make a determination within 4 labels: *Negative for Pneumonia (NP)*, *Typical Appearance (TA)*, *Indeterminate Appearance (IA)* and *Atypical Appearance (AA)*. For each test image in detection task, a bounding box along with the *Opacity* class for findings and a corresponding confident score should be predicted. Otherwise, if there are no findings, the prediction should be *none 1 0 0 1 1* (*none* is the class ID for no finding, and this provides a one-pixel bounding box with a confidence of 1.0). We aim to accelerate the performance by updating the latest classifiers and detectors on huge datasets and further comparing with other researches in both performance and precision. Besides, many advanced techniques and tricks such as image augmentation, ensemble learning are implemented.

3 Proposed Framework

This research proposes a framework for both classification and detection of COVID-19 abnormalities as illustrated in Fig. 1. For classification task, an input CXR which tagged as *study level* is inferred by the 4-classes classifiers. Meanwhile, the detection task is applied to CXRs marked as *image level*. At first, the input image is categorized into *Finding* or *No finding* by an EfficientNetV2 pre-trained as a 2-classes classifier to eliminate false positives. The *No finding* class indicates that there is no abnormality in this image, hence the ultimate prediction is *None*. On the contrary, the *Finding* sample is passed through a lung detector. Next, the dissected thorax is inferred by three detectors to find the location of the lesions which is labelled as *Opacity*. Finally, predicted bounding boxes and corresponding confident scores are then ensemble using weighted box fusion method to produce location of the lesion.

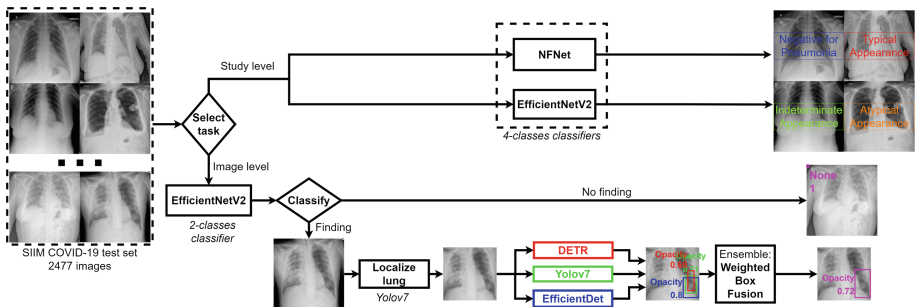


Fig. 1. Our framework for COVID-19 abnormalities classification and detection.

The training process of the models utilized in the framework is depicted in Fig. 2. The SIIM COVID-19 dataset is the primary dataset used in this study for training and assessing models. To improve the generalization of the models, we also gather various external datasets. All data is preprocessed according to the method described in Subsect. 4.1, and image augmentation is performed as outlined in Subsect. 3.1. The models are trained through multiple iterations, with the best models selected through k-fold cross validation. These top models are then consolidated into a model zoo for use in the inference phase.

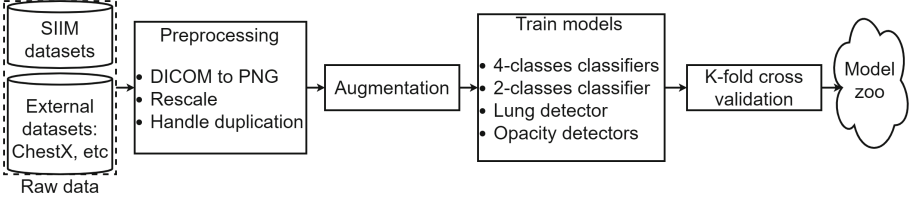


Fig. 2. Flow of training and validating models from raw data.

3.1 Data Augmentation

Image augmentation is one of the key techniques used in deep learning and computer vision tasks to increase the quality and generality of trained models by creating fresh training samples from the existing data. The most prevalent augmentation strategies according to the expertise in medical image analysis are Blur, CLAHE, Cutout, ElasticTransform, GridDistortion, HorizontalFlip, IAASharpener, PiecewiseAffine, RandomResizedCrop, ShiftScaleRotate, and VerticalFlip. In this work, we adopted all of these method through the Albumentation library [2] while training models in the both tasks. Additionally, we apply Mosaic augmentation for detection task since this method not only modifies data on one image but also randomly combines data from four images into a new synthetic one as depicted in Fig. 3.

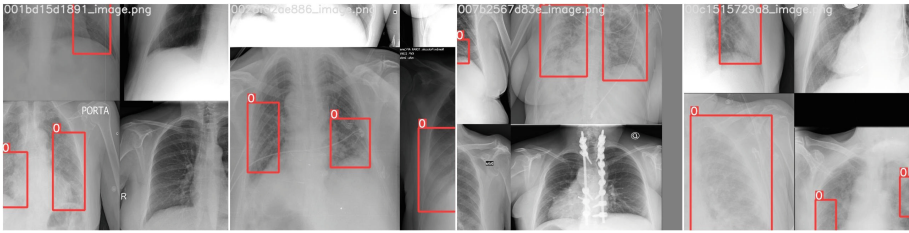


Fig. 3. Synthetic labels generated by Mosaic augmentation

3.2 COVID-19 Abnormalities Classifier

For classification task, two separate architectures, Normalizer-Free Networks (NFNet) [1] and EfficientNetV2 [17], are trained as 4-classes classifier to classify input image into 4 categories: *NP*, *TA*, *IA* and *AA*. NFNet is designed with the

aim of overcoming instabilities for large learning rate or strong data augmentation with the proposal of the adaptive gradient clipping, which clips gradients based on the unit-wise ratio of gradient norms to parameter norms. A family of Normalizer-Free ResNets from NFNet-F0 to NFNet-F6 are constructed and achieved higher validation accuracies on ImageNet while being faster to train. In this work, a NFNet-F5 is firstly pretrained on the CheXpert [8] and MIMIC-CXR [9] dataset and then fine-tuned on the main dataset at the image size of 512×512 . EfficientNetV2 optimizes training time and parameter efficiency by combining training-aware neural architecture search and scaling. Crucial characteristics of the EfficientNetV2's backbone are: it heavily uses both MBConv and the newly introduced Fused-MBConv in the initial layers; it favors a lower kernel size but adds more layers; it eliminates the final stride-1 state employed in EfficientNet. During training, the regularization and image size are also adaptively changed as the concept of progressive learning to assist the network learn basic representations quickly and simply. Transfer learning technique are applied in this research on an EfficientNetV2-M which is pre-trained on the ChestX-ray14 dataset [23] with the aim of making it more general. Both of the two classifiers use a Binary Cross Entropy loss $\mathcal{L}_{BCE}$ blended loss of each class as in Eq. 1. The hyperparameters are selected empirically as follow: $\alpha = 0.2$, $\beta = 0.2$ and $\gamma = 0.3$ in order to reduce class imbalance phenomenon. Experiments show that this contrivance improves both cross-validation as well as final submission score.

$$\mathcal{L}_{BCE} = \alpha \mathcal{L}_{negative} + \beta \mathcal{L}_{typical} + \gamma \mathcal{L}_{indeterminate} + (1 - \alpha - \beta - \gamma) \mathcal{L}_{atypical} \quad (1)$$

For detection task, instead of feeding chest radiographs directly to the object detectors, we first use a 2-classes classifier to filter out images that may contain abnormalities associated with the COVID-19 disease. As shown in Fig. 1, a EfficientNetV2-S is chosen as a classifier to separate inputs CXR into two categories *No finding* and *Finding*, corresponding to normal image and abnormal image respectively. This model is trained on these two classes extracted from the main dataset where we process *Negative for Pneumonia* labels as *No finding* and other labels as *Finding*. Initial weights are also pre-trained on the RICORD dataset [19] which contains 1000 CXRs rated by 3 different radiologists using the same 4-classes taxonomy as in the main dataset. This 2-classes classifier scheme brings two benefits: reduction of false positive rate and acceleration of speed during the inference phase. Because classifiers are frequently more accurate than detectors, this technique is extremely effective in filtering out early negative samples, especially for imbalanced datasets as observed in similar studies [13]. This rapid classifier also narrows down the CXRs that must be inferred on by the slower object detection models. For example, in the training dataset, there are 1736 *No finding* images out of 6334. This implies that more than a fourth of the entire data are eliminated, which leads to speed acceleration.

3.3 Lung Detector

The purpose of lung detection is to separate between voxels that correspond to lung tissue and those that relate to surrounding anatomy. This localization

serves to reduce background noise to assist the posterior detectors focus more on lung lesions, based on the assumption that nCoV virus-associated lesions are exclusively found in the lung area. A typical method for this task is the simple center-crop at about 70% to 80% of the image. This approach proves to be effective and robust in cases where the lung is exactly in the center of the image. However, several factors influence this, including the capture angle, the radiologist’s alignment, and the patient’s standing position. When we examine the samples in this work’s dataset, we see that the rib cage frequently deviates from the center of the picture. Therefore, a Yolov7 architecture is chosen instead of the center-crop method. 6334 CXRs in the training dataset are manually annotated to extract the lung bounding box. This data are then fed into training loops where a Yolov7 is employed to train on a single fold. Experiments show that this model boosts average precision while comparing with the simple center-crop.

3.4 COVID-19 Lesion Detector

The goal is to localize all findings, of the unique class *Opacity*, on a lung-cropped CXR with bounding boxes along with a confidence score. An example is shown as in Fig. 1 in which SOTA detectors (DETR, Yolov7 and EfficientDet) predict three different bounding boxes with score of 0.72, 0.65 and 0.78 respectively. DETR [3], combines a transformer encoder-decoder architecture with a set-based global loss that requires bi-partite matching to force unique predictions. Yolov7 [21] focuses on improving the accuracy without increasing the inference cost. EfficientDet [18] proposes a weighted bi-directional feature pyramid network for multiscale feature fusion and a compound scaling approach.

For implementation, DETR models are trained with ResNet-101 as backbone, an encoder-decoder transformer, a feed forward network and a learning rate of 0.001. Yolov7-E6 models are trained with E-ELAN as backbone and a learning rate of 0.005. EfficientDet-D7 model are also trained with SGD optimizer with a learning rate of 0.001 and EfficientNet as backbone. A Cosine Annealing scheduler and a Focal Loss function are applied. All backbone networks are pretrained on the RSNA dataset with *Opacity* class. Input image are scaled to 1024×1024 with a random augmentation described in Sect. 3.1 in training while during inference time, test time augmentation only consists of the horizontal and vertical flip. K-fold cross validation is employed to identify problems such as overfitting or selection bias, as well as to provide insight into how the models will generalize to an independent dataset, by splitting the training dataset into 5 folds based on patient’s ID metadata extracted from DICOM data in the training stage.

3.5 Ensemble Technique

The ensemble technique combines predictions of several models to achieve higher predictive performance than each of the constituent one could accomplish alone. These predictions consist of coordinates of the bounding box rectangle selected using non-maximum suppression based methods [11]. However, these approaches only produce boxes from a single model and can not adequately yield blended localization of predictions combined from various models. Therefore, we adopt

the Weighted Boxes Fusion [16] in which coordinated of bounding rectangles as well as confidence scores from models are fused to construct the average boxes.

4 Experiments

4.1 Data Collection and Evaluation Protocol

Our proposed method is evaluated on different benchmark datasets. Table 1 summarizes data collection including 7 available CXR datasets from public sources. They are improved by augmentation techniques mentioned in Subsect. 3.1. In this work, a NFNet-F5 is firstly pretrained on the CheXpert dataset [8] and then fine-tuned on the main dataset at the image size of 512×512 . Transfer learning technique are applied in this research on an EfficienNetV2-M which is pre-trained on the ChestX-ray14 dataset [23] in order to improve its generalization. Initial weights are also pre-trained on the RICORD dataset [19] which contains 1000 CXRs rated by 3 different radiologists using the same 4-classes taxonomy as in the main dataset. The data is split in the ratio of 80:20 for the training and testing subsets. All images are in DICOM format and have been de-identified to safeguard patient privacy. A panel of qualified radiologists labels them based on the presence of opacity as well as the overall appearance.

Table 1. Description of collected CXR datasets.

Dataset	#Observations	#Image	#Patient	Annotation	Year
ChestX-ray14 [23]	15: No finding, 14 Pathologies	112,120	30,805	Classification	2017
MIMIC-CXR [9]	14: No finding, 12 Pathologies, Support devices	377,110	65,379	Classification	2019
CheXpert [8]	14: No finding, 12 Pathologies, Support devices	224,316	65,240	Classification	2019
VinDr-CXR [12]	28: No finding, 22 Lesions, 5 Diseases	18,000	18000	Detection	2020
BIMCV COVID-19+ [20]	20: Normal, COVID-19, Other findings	5381	1311	Classification	2020
RICORD [19]	4: Typical, Indeterminate, Atypical, Negative appearance for COVID-19	1,000	361	Classification	2021
SIIM [10]	4: Typical, Indeterminate, Atypical, Negative appearance for COVID-19 2: No finding, Opacity	7897	7268	Classification Detection	2021

All experiments are conducted on a RTX 1080. F1-score and mean Average Precision (mAP) at IoU 0.5:0.95 are employed as evaluation metrics.

4.2 Results and Discussions

The experimental results of the classifiers are shown in Table 2. Covid-net [22] is selected as baseline model for comparing with the classifiers in this research. We modified the last fully connected layer and retrain the model to adapt with our datasets. Experiments show that the baseline model proved to be inefficient with 2 classes IA and AA, which are unavailable in it's original dataset. Our classifiers generally achieved better performance on the testing dataset.

The outcomes reveal a significant disparity between the four classes. When inferring the NP class, the EfficientNetv2 model earns the greatest F1 score

Table 2. F1 score of the COVID-19 abnormalities classifiers on the SIIM test set

Classifier	F1 score			
	Negative for pneumonia	Typical appearance	Indeterminate appearance	Atypical appearance
Covid-net [22]	0.71	0.74	0.48	0.53
NFNet [1]	0.79	0.72	0.59	0.56
EfficientNetv2 [17]	0.83	0.75	0.61	0.54

of 0.83, while the NFNet model meets 0.79. The TA class achieved the second highest results, at 0.75 and 0.72 respectively for EfficientNetv2 and NFNet. The lowest result belongs to the case of the AA class, only 0.56 and 0.54 for the two classifiers. The challenges come from the complexity of the “Atypical” class definition, which includes a range of abnormalities reported uncommonly for COVID-19 pneumonia such as pneumothorax or pleural effusion, pulmonary edema, lobar fusion, solitary pulmonary mass or nodule, diffuse small nodules and cavity. The model might not have enough samples of each abnormality to draw useful representations from given the variety of ways that unusual things might appear along with the dearth of data available for this class (10%). The features of the IA layer are also unclear, which leads to its low result, 0.59 and 0.61, respectively. There is a huge difference between the results of 2 classes NP and TA compared to 2 classes AA and IA. This is reasonable because symptoms on the NP and TA classes are often very specific and easily distinguishable, representing virus-positive and negative cases, respectively. In addition, the number of samples of these two classes also accounts for the majority of the test set.

Table 3 shows comparison between methods integrated in the proposed framework. If only the discrete models are counted, Yolov7 clearly outperforms the other two models with the best score of 0.563. A solid outcome was also achieved by the DETR architecture, which reached 0.542 accuracy points. When the EfficientDet model only managed to attain an accuracy of about 0.5, which is roughly 10% less than DETR and Yolov7, it became clear that it was not appropriate for this kind of situation. This can be accounted for by the lack of ample and diverse training data for a data-hungry architecture like EfficientNet.

Table 3. Evaluation of the COVID-19 lesion detector on the SIIM test set

Detector	Accuracy (mAP@ IoU 0.5:0.95)		Performance		
	Discrete model	Lung localized	Inference speed (FPS)	GPU memory requirement (MB)	Model size (MB)
DETR [3]	0.542	0.587	25	2683	232
Yolov7 [21]	0.563	0.591	34	3520	290
EfficientDet [18]	0.499	0.574	19	1903	187
Our Fusion	0.605	0.612	8	8106	709

The lung region identification and separation strategy proved to be very effective as it improved the accuracy for all models. For each of the three situations, the accuracy increased clearly, by 5%, 8%, and 15%, respectively. The most noteworthy of which is the application of lung separation by Yolov7 followed by the

use of EfficientDet, which has produced a result of 0.499 to 0.591, an increase of 15%. The reason for this is most likely because the EfficientDet model, with the help of the lung partition model, has made it easier to focus on the core region of the lung. The case of overlapping two YOLOv7 models proved to be ineffective when the results only increased by about 5%, this is because the architectural repetition in the two phases reduces the diversity of the results.

It should be noticed that the accuracy is significantly increased by employing the weighted box fusion technique, which merges predictions from three different models. It achieves an accuracy of up to 0.605, which is 7% higher than YOLOv7 and a 21% increase over EfficientDet. The fusion solution reached an outstanding score of 0.612 in the case of previously recognized lung regions. However, as compared to the single findings, this outcome is only moderately improved by roughly 4%. This can be explained that the accuracy of the alternatives prior to compositing is already fairly high, so blending the results makes little sense.

Concerning performance evaluation, three factors are addressed. For inference speed, the YOLOv7 architecture achieved the best results, 34 fps. DETR comes in second with 25 fps and EfficientDet comes in third with 19 fps. When all three models were performed in succession, the calculation speed dropped considerably to only 8 fps. In terms of GPU memory required when referencing the model, the best result belongs to EfficientDet with about 2 GB of GPU, while DETR (resp. YOLOv7) consumes about 2.6GB (resp. 3.5 GB). If performing parallel models, the required memory is up to more than 8 GB. For memory space, EfficientDet needs only 187 MB, while DETR (resp. YOLOv7) takes up 232 MB (resp. 290 MB).

The preceding results are used to make recommendations for the implementation of deep learning models in real-world scenarios. Diagnostic errors need to be kept to a minimum in both the general medical profession and COVID detection in particular. In order to merge these three models in a way that provides the best level of accuracy, it is important to employ the lung region separation technique and the weighted box fusion approach. This manner is much more expensive, it often results in a reduction of more than 4 times the computation speed when referring the model, increasing both the amount of GPU memory needed and the size of the storage. This approach is appropriate in cases where the accuracy is a priority, such as severe or highly suspected cases. On the other hand, we can use the YOLOv7 straightforward model for a better trade-off between the accuracy and the consuming time. Though there may be a little, 5%, reduction in precision, there is a four-fold gain in calculating speed. The EfficientDet model can be used when the available hardware is constrained in terms of capacity because it consumes small resources to achieve satisfactory accuracy. Alternately, we can adopt the DETR model, which will provide a greater precision while only slightly increasing the amount of resources needed.

5 Conclusions

We have proposed a framework consisting of two classifiers and three detectors enhanced by ensemble learning approach to categorize and locate COVID-19

related anomalies. The results are thoroughly analyzed to provide recommendations in the application of deep learning models in practice. The limitations in data imbalance as well as the performance of the models will be improved in future researches. Our code is available at this [address](#) for research purpose.

References

1. Brock, A., De, S., Smith, S.L., Simonyan, K.: High-performance large-scale image recognition without normalization. arXiv preprint [arXiv:2102.06171](#) (2021)
2. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, M., Kalinin, A.A.: Albumentations: fast and flexible image augmentations. *Information* **11**(2) (2020)
3. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers (2020)
4. Cohen, J.P., Morrison, P., Dao, L.: Covid-19 image data collection (2020). <https://doi.org/10.48550/ARXIV.2003.11597>. <https://arxiv.org/abs/2003.11597>
5. Costa, Y.M.G., et al.: Covid-19 detection on chest x-ray and ct scan: a review of the top-100 most cited papers. *Sensors* **22**(19) (2022)
6. Gomes, R., Kamrowski, C., Langlois, J., C., Haley, M.: A comprehensive review of machine learning used to combat COVID-19. *Diagnostics (Basel)*, p. 1853 (2022)
7. Huang, C., et al.: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**(10223), 497–506 (2020)
8. Irvin, J., Rajpurkar, P., Ilcus, S.: Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. *CoRR abs/1901.07031* (2019)
9. Johnson, A.E.W., et al.: MIMIC-CXR: a large publicly available database of labeled chest radiographs. *CoRR abs/1901.07042* (2019)
10. Lakhani, P., et al.: The 2021 siim-fisabio-rsna machine learning covid-19 challenge: annotation and standard exam classification of covid-19 chest radiographs, October 2021
11. Neubeck, A., Van Gool, L.: Efficient non-maximum suppression. In: 18th International Conference on Pattern Recognition (ICPR 2006), vol. 3, pp. 850–855 (2006)
12. Nguyen, H.Q., et al.: Vindr-cxr: an open dataset of chest x-rays with radiologist's annotations (2020)
13. Pham, V.T., Tran, C.M., Zheng, S., Vu, T.M., Nath, S.: Chest x-ray abnormalities localization via ensemble of deep convolutional neural networks. In: ATC, pp. 125–130 (2021)
14. Rubin, G.D., Ryerson, C.J.: The role of chest imaging in patient management during the covid-19 pandemic: a multinational consensus statement from the fleischner society. *Radiology* **296**(1), 172–180 (2020)
15. Shi, F., et al.: Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for covid-19. *IEEE Rev. Biomed. Eng.* **14**, 4–15 (2021)
16. Solovyev, R.A., Wang, W.: Weighted boxes fusion: ensembling boxes for object detection models. *CoRR abs/1910.13302* (2019)
17. Tan, M., Le, Q.V.: Efficientnetv2: smaller models and faster training. *CoRR abs/2104.00298* (2021). <https://arxiv.org/abs/2104.00298>
18. Tan, M., Pang, R., Le, Q.V.: Efficientdet: scalable and efficient object detection. *CoRR abs/1911.09070* (2019). <http://arxiv.org/abs/1911.09070>
19. Tsai, E., Simpson: The rsna international covid-19 open radiology database(ricord). *Radiology* **299**(1), E204–E213 (2021)

20. Vayá, M.d.l.I., et al.: Bimcv covid-19+: a large annotated dataset of rx and ct images from covid-19 patients (2020)
21. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors (2022)
22. Wang, L., Lin, Z.Q., Wong, A.: Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Sci. Rep.* **10**(1), 1–12 (2020)
23. Wang, X., Peng, Y., Lu, L., Bagheri, M., Summers, R.M.: Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *CoRR* abs/1705.02315 (2017)



Breathing Pattern Assessment Through the Empirical Mode Decomposition and the Empirical Wavelet Transform Algorithms

Zakaria El Khadiri^(✉), Rachid Latif, and Amine Saddik

Laboratory of Systems Engineering and Information Technology LiSTi, National School of Applied Sciences, Ibn Zohr University Agadir, Agadir, Morocco
`zakaria.elkhadiri@edu.uiz.ac.ma`

Abstract. In healthcare applications, the photoplethysmography concept is widely employed to estimate the heart rate, breathing rate, and other vital signs of humans, by using facial videos. However, the latter physiological indicators are fundamentally used to determine the physiological and pathological state of a person. In consequence, the present study will notably concentrate on the respiratory rate value extraction through a non-contact technique. The current approach is mainly based on image processing for PPG data gathering, and plethysmography (PPG) signal normalization, signal decomposition, and Fourier transform in order to extract the adequate candidate for breathing rate estimation. Moreover, our study contains a result comparison of the decomposition eminent methods such as the Empirical Mode Decomposition (EMD) and Empirical Wavelet Transform (EWT), the metrological interpretation of the results achieved from the experiment's research is discussed and conclusions are presented.

Keywords: Photoplethysmographic · Respiratory rate · EMD · EWT · Fourier transform

1 Introduction

Vital signs are measurements of the body's most basic functions, however, they have a very role to detect and track medical problems and to routinely checked by healthcare providers. Most of these physiological parameters are body temperature, pulse rate, breathing rate, and blood pressure, while they can be measured in a medical setting, at home, at the site of a medical emergency, and among others. Minutely, the respiratory rate is a non-invasive physiological sign and a helpful evaluation tool to indicate patient deterioration and degradation [1], also it presents the number of breaths per minute and expresses the breathing rate in an organism. This metabolic process provisions the necessary oxygen to the

Supported by organization Laboratory of Systems Engineering and Information Technology LiSTi.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
A. E. Hassanien et al. (Eds.): AICV 2023, LNDECT 164, pp. 262–271, 2023.
https://doi.org/10.1007/978-3-031-27762-7_25

whole cells of the human body, and deportee out the carbon dioxide. Although, the normal breathing rate is counted between 12 and 20 respiration per minute. While a respiration rate under 12 or over 25 breaths per minute is considered abnormal.

Signal decomposition, denoising, and decompression fields are a set of the most important recent processed topics. More recently, several researchers have presented a review study of the common methods used for the same purpose. On our side, our main focus will be notably shed on the signal decomposition signals techniques which serve to extract the different intrinsic modes of oscillation in conditions to have the same number of extrema and zero-crossing. Moreover, the more challenging approach for all decomposition purposes is the single component separate from the mixed signals. Where the composite signal consists of a sample-wise superposition from multiple components.

In 2020, R. Lei et al. [2] proposes a framework combining the complementary ensemble empirical mode decomposition technique with both the independent component analysis and the non-negative matrix factorization for heart rate and respiratory rate estimation from the photoplethysmography (PPG) signal. The authors have achieved a significant result that outperforms both the digital filtering approach and the conventional empirical mode decomposition-based methods with higher accuracy and reliability for physiological indicators estimating. Additionally, H. Sharma [3], suggests a simple effective approach for accurate estimation of respiratory rate (RR) from the photoplethysmogram (PPG) based on the ensemble empirical mode decomposition (EEMD) and Kalman filter coupled. The authors have employed two independent datasets, namely CapnoBase and MIMIC, and they have achieved a lower error during the estimation.

In 2022, T. Iqbal et al. [4] proposes a new algorithm to extract the breathing rate through a photoplethysmography sensor signal for health monitoring. This study shows the well-deal with the low-quality signals retrieved from the sensors and removes the undesired noises. They have achieved an important accuracy level and robustness against noises by unifying selective windowing, preprocessing, and signal conditioning, modified Welch filtering, and postprocessing steps. In the end, they evaluated the performance of the proposed method using the PPG data existing in the public dataset that's called Berth Israel Deaconess Medical Centre (BIDMC) which was benchmarked against different existing methods utilizing several performance evaluation metrics indicators.

The remainder of this paper has been orderly in different sections. Section 2 introduces the most common algorithms involved to gather the instantaneous breathing rate values whichever is linked to image processing, or signal processing methods. Section 3 details the stacking layers of our proposed system for respiratory rate monitoring. Section 4 presents the obtained results from the proposed system and their interpretation and explanation. Finally, we provide a brief conclusion and future directions in Sect. 5.

2 Algorithms

As abovementioned, the current pattern implementation is founded on two notable processes. The first one is mainly served to manipulate and process the sequence received image from such as webcam, the latter system will be titled “face’s PPG data extraction”. While, the second part will mainly help to deal with the output signal obtained from the first system, to estimate the respiratory rate values, the latter system will be titled “respiratory rate estimation”. For this reason, our suggested approach will be primarily established on the following commons algorithms.

2.1 Face’s PPG Data Extraction

Face’s data extraction is the first part of the entire proposed system for respiratory rate estimation. This block is considered the finer part to provide a first interpretation of the plethysmography signal (PPG). However, on our side, we will be based on the famous object detection technique which is termed the Viola and Jones [5] in order to detect and track the region of interest of the human’s face. The latter method has reached a significant accuracy in the object detector field.

2.2 Respiratory Rate Estimation

Empirical Mode Decomposition. EMD, Empirical Mode Decomposition, is a signal decomposition method utilized to decompose the input signal into various constituent signal components termed Intrinsic Mode Functions (IMFs) [6]. The latter decomposition concerns the recursively stationary and non-stationary signals based on their frequency and amplitude at a certain time. Overall, the EMD is most well-used for physiological indicators estimating from biological signals and for eliminating the PPG signal’s interferences. On a mathematical hand, the relationship between the original $x(t)$ signal to the IMFs and its residue is quite detailed as follows:

$$x(t) = \sum_{i=1}^N IMF_i(t) + r(t) \quad (1)$$

Based on the above equation, the steps of the EMD algorithm can be expressed as follows:

1. Determining local minima and maxima of the input signal $x(t)$.
2. Extracting the lower envelope of the signal by interpolating all local minima and the same for all local maxima to extract the upper envelope as well.
3. Compute the mean of the local maxima and minima, the mid-value will be termed $m(t)$.
4. Computing the candidates of IMFs: $IMF(t) = x(t) - m(t)$.
5. Based on the below criteria, we can conclude the IMFs:

- 5.1 The number of zero crossing and extrema points should be the same or different by 1.
- 5.2 The average envelope should be zero at all points.
6. The remaining IMFs can be extracted by repeating the steps from 1 to 5.
7. Producing the residue $r(t)$ signal with monotonous values.

The outstanding inconvenience of the EMD approach is mostly located in the mathematical theory lack that still subsisted. Nevertheless, it has achieved a great number of successful applications in several domains.

Empirical Wavelet Transform. EWT is an adaptive signal analysis technique that served to decompose the non-stationary and non-linear given signals into a series of different components and gather the finer features from the obtained modes. This method has been proposed in 2013 by J. Gilles et al. [7] and is mainly inspired from the Empirical Mode Decomposition (EMD) and Wavelet Transform (WT) [8] methods. EWT aims notably with a consolidated mathematical theory in order to tackle EMD's limitations and to deal with the known shortcomings which greatly limit EMD's limitations. Minutely, the below flow chart is clearly detailed the EWT algorithm (Fig. 1).

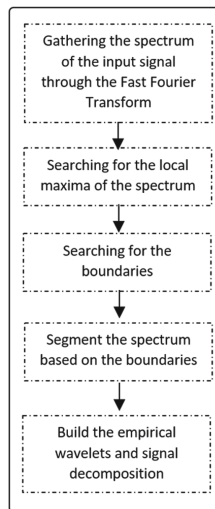


Fig. 1. EWT method

Fast Fourier Transform. Fast Fourier Transform, is a method that calculates the Discrete Fourier Transform (DFT) of a sequence, in the mathematical side, Fourier analysis translates a signal $x(t)$ from its original domain (usually a

time domain) into a frequency domain $X(\omega)$ and vice versa. The transformation equation can be expressed as follows:

$$X(\omega) = \int_{-\infty}^{+\infty} x(t)e^{-i\omega t} dt \quad (2)$$

However, the FFT can be considered the fast computational method for a DFT, which means, producing an array of frequency-domain spectrum samples based on an array of a time-domain wave. The N samples of the input array must be a power of 2 for the FFT application with all real values, then, the algorithm will generate a complex-valued sample [9,10].

3 Proposed System

As aforementioned, our suggested system is devoted to extracting the breathing rate value from facial image frames using a digital RGB camera or a webcam. Therefore, the current proposed architecture is mainly based on several techniques widely employed by the image and signal processing community. However, those methods would help to estimate some of the major physiological indicators such as heart rate (HR), respiratory rate (RR), and heart rate variability (HRV) to monitor a person's state. On our side, after gathering the necessary PPG data from human faces using an RGB camera, our system will be based only on the green band. Indeed, the main reason of green channel is linked to the normalized PPG signal will be mainly subjected to decomposition using two famous approaches which are: the Empirical Mode Decomposition (EMD) and the Empirical Wavelet Transform (EWT) (Fig. 2).

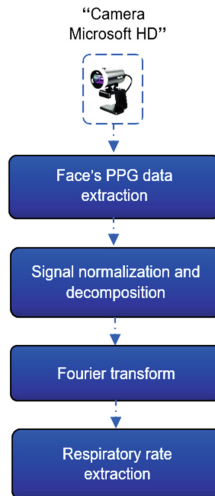


Fig. 2. Proposed system

3.1 Face's PPG Data Extraction

As depicted above, our first main block notably served to process the successive frames collected from the webcam using the famous object detection framework to track human face areas. Afterward, based only on the green channel, the latter choice was inspired from the works proposed by M. Poh, and W. Verkruyse, [11, 12] (Fig. 3).

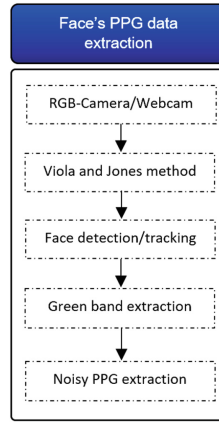


Fig. 3. Face's PPG data extraction block

3.2 Signal Normalization and Decomposition

Signal normalization and decomposition is the second main block of the entire proposed system, however, it occupied the major part and is considered the prominent function to provide the relevant signal that contains the breathing data information. For more convenience, Fig. 4 below depicts the stacking functional block involved to extract respiratory rate value.

4 Results and Discussion

As depicted above, the proposed work is mainly dedicated to estimating the breathing rate from facial image sequences using a plethysmography signal analysis, however, after assessing each frame extracted from a webcam and extracting the first noisy normalized PPG signal, the latter will be subjected to signal decomposition. This extraction is performed via C/C++ and Python tools and the OpenCV library. In the first stage, the results of the PPG decomposition with the Empirical Mode Decomposition are clearly shown in Fig. 5 below.

Figure 5 shows the Intrinsic Mode Functions extracted from the green band after applying the EMD technique. As clearly shown, the IMFs components have

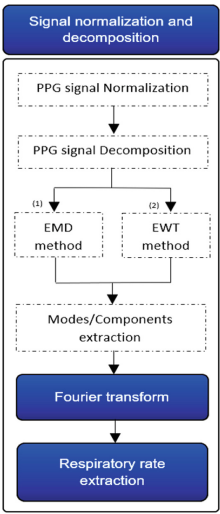


Fig. 4. Signal normalization and decomposition block

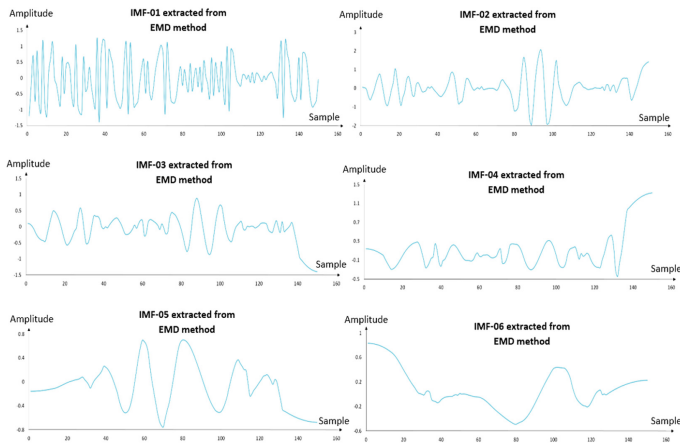


Fig. 5. Curves extracted from the EMD approach

been obtained after applying the Hilbert-Huang Transformation which they took an oscillatory part of the signal format. At the same time, the normalized PPG signal will be processed by the second decomposition signal function (EWT), for this case, Fig. 6 shows the principal components extracted through the Empirical Wavelet Transform method:

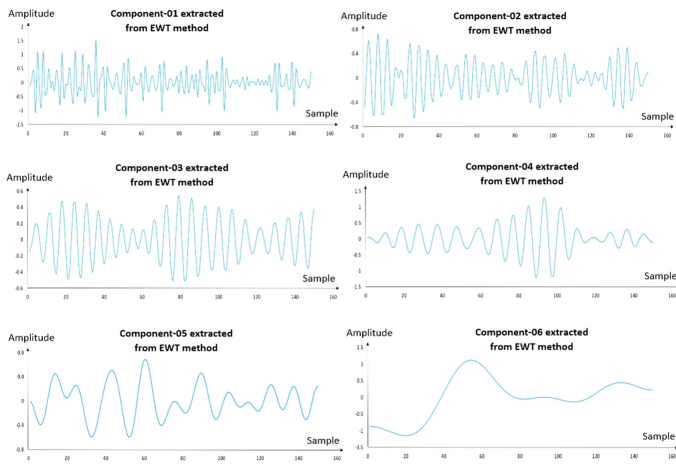


Fig. 6. Curves extracted from the EWT approach

After decomposing the PPG signal into several IMFs, a Fourier Transform analysis will be applied as well to define which component is the relevant one for respiration rate information. However, the based criterion is that the mode must contain a spectral power dominance at frequencies 1 Hz, while the remaining components which have a frequency more 1 Hz are not considered to be a candidate for respiratory rate estimation, whilst, they could be more correspond to the heart rate or noises. Afterward, the breathing rate value can be easily estimated through the maximum power response index. Additionally, our experiment’s research was validated based on a database of 10 samples to show the correlation and closeness between both methods and to compare their alignment, however, our outcome reveals that the breathing rate estimation using EWT seems more accurate compared with the EMD with a standard deviation value equal 0.643 rates Vs 1.199 achieved using EMD. The following Table 1 and Fig. 7 illustrate the mean and the standard deviation of respiratory rate value and its grouped histogram.

According to the above results, we concluded the outperforming consistent results of EWT over EMD either in breathing rate estimation or in terms of execution time which can provide the estimated modes faster than EMD, this finding will be explained due to the lack of mathematical background involved in the EMD concept which leads to an overestimating of the number of modes. Even though, in order to respond to real-time constraints, both methods are actually impractical to be used such as a system that doesn’t contains a big computer performance.

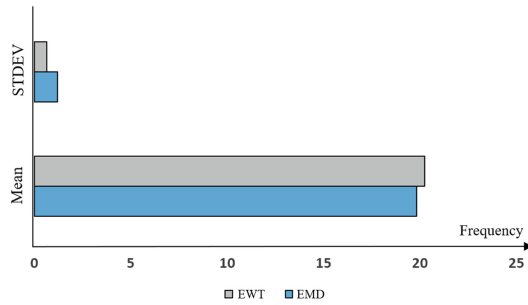


Fig. 7. Grouped histogram Mean and standard deviation

Table 1. Mean and standard deviation values

	EMD	EWT
Mean	19.84 beats/min	20.25 beats/min
Standard deviation	1.199 beats/min	0.643 beats/min

5 Conclusion

This paper describes a comparison between two main techniques largely utilized for signal decomposition field which are Empirical Mode Decomposition and Empirical Wavelet Transform to determine the physiological indicators especially the respiratory rate, the current simulation was carried out using MATLAB and Python programming languages. At the end of this research, the used methods can be the basis for determining additional vital parameters such as heart rate (HR), arterial blood oxygen saturation, and HRV, among other. Overall, the proposed system can also be improved by applying additional image pre-processing algorithms and further signal filtering to improve the quality of the face’s frame and reduce the unexpected noises from the first PPG noisy signal.

References

1. Rolfe, S.: The importance of respiratory rate monitoring. *Br. J. Nurs.* **28**(8), 504–508 (2020)
2. Lei, R., Ling, B.W.K., Feng, P., Chen, J.: Estimation of heart rate and respiratory rate from PPG signal using complementary ensemble empirical mode decomposition with both independent component analysis and non-negative matrix factorization. *Sensors* **20**(11), 3238 (2020)
3. Sharma, H.: Extraction of respiratory rate from PPG using ensemble empirical mode decomposition with Kalman filter. *Electron. Lett.* **56**(13), 650–653 (2020)
4. Iqbal, T., Elahi, A., Ganly, S., et al.: Photoplethysmography-based respiratory rate estimation algorithm for health monitoring applications. *J. Med. Biol. Eng.* **42**(2), 242–252 (2022)

5. Viola, P., Jones, M.: Robust real-time face detection. In: Eighth IEEE International Conference on Computer Vision, ICCV 2001. Proceedings, pp. 747–747 (2001)
6. Huang, N.E., et al.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Roy. Soc. London Ser. A Math. Phys. Eng. Sci.* **454**(1971), 903–995 (1998)
7. Gilles, J.: Empirical wavelet transform. *IEEE Trans. Signal Process.* **61**(16), 3999–4010 (2013)
8. Morlet, J., Arens, G., Fourgeau, E., Giard, D.: Wave propagation and sampling theory-Part I: complex signal and scattering in multilayered media. *Geophysics* **47**(2), 203–221 (1982)
9. Heckbert, P.: Fourier Transforms and the Fast Fourier Transform (FFT) algorithm. *Notes Comput. Graph.* **3**, 15–463 (1995)
10. Wörner, S.: Fast Fourier transform numerical analysis seminar (2008)
11. Poh, M.Z., McDuff, D.J., Picard, R.W.: Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express* **18**(10), 10762–10774 (2010)
12. Verkruyse, W., Svaasand, L.O., Nelson, J.S.: Remote plethysmographic imaging using ambient light. *Opt. Express* **16**(26), 21434–21445 (2008)



Lung Cancer Stages Classification Based on Differential Gene Expression

Moshira S. Ghaleb^(✉), Hala M. Ebied, and Mohamed F. Tolba

Scientific Computing Department Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

{Moshirasg, halam, fahmytolba}@cis.asu.edu.eg

Abstract. A form of cancer known as lung cancer usually appears in the cells that line the lungs' air passages. It is the primary cause of cancer-related deaths in both genders. Lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) are two separate types of non-small cell lung cancer (NSCLC). Recognition of the differentially expressed genes (DEGs) is necessary to distinguish between healthy and diseased states molecularly. Biomedical books and microarray investigations are two common places to find details concerning gene expression variations. A gene is said to be highly expressed if there is a significant statistical distinction or change in reading counts or expression level indices between two different experimental conditions. A good gene expression strategy leads to a good gene silencing strategy, which subsequently slows or stops the cancer progression. In this paper, we present a robust analysis of the mRNA data from the LUAD data to find the most differentially expressed genes. Then we employ several machine learning techniques to categorize the samples into the five stages of lung cancer. We investigated four distinct models; including linear regression, logistic regression, RANSAC regression, and decision trees. The models achieved 100%, 91%, 98%, and 100% accuracies for 12 differential expression genes using linear regression, logistic regression, RANSAC regression, and decision trees respectively. The logistic regression and decision trees achieved 100% and 98% accuracies for 5665 differential expression genes. The proposed models achieved the highest accuracies in comparison with the state-of-the-art models.

Keywords: Gene expression · DEGs · LUAD · Lung cancer · Classification · Machine learning · Decision tree · RANSAC · Logistic regression · Linear regression

1 Introduction

The second-most fatal illness in the world today, behind cardiovascular disorders, is cancer. Over the past few decades, enthusiasm for delivering accurate diagnosis tools has increased. When certain cells behave abnormally and divide uncontrollably, they harm adjacent cells and form a mass or tumor that, in some circumstances, can be fatal [1]. The likelihood of harm to other cells can be decreased by early detection and appropriate therapy. The high prevalence of cancer-related mortality [2] is driving scientists to create

innovative systems for the early detection of cancer and categorization. However, because cancer cells are disorganized, early diagnosis is particularly challenging. This is where RNA-Seq analysis is quite beneficial.

Small-cell lung carcinoma (SCLC), which accounts for around 15–20% of lung cancer cases, and non-small-cell lung carcinoma (NSCLC), which accounts for about 80–85% of lung cancer cases [4], are the two primary forms that can be distinguished within them [3]. Two distinct kinds of NSCLC—adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC)—can be distinguished. LUAD typically originates from peripheral lung tissue and is linked to nonsmokers who have never smoked [5].

By delivering more normalized and less misleading information for prediction and categorization objectives, RNA-Seq is a recent and well-liked approach that is utilized to find new transcripts and isoforms [6, 7]. Transcriptase profiling's most important role is to identify genes that are differentially expressed in the body or to find variations in genes at various levels [8]. RNA sequencing can be used to identify and quantify RNA in one location [9]. The classification of diseases such as breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), renal homophobe, etc. is done using RNA-Seq data, which are freely accessible from many databases [10].

The understanding of phenotypic variation depends on the accurate recognition of differentially expressed genes (DEGs) between particular conditions. In order to identify the gene that causes diseases most effectively, RNA gene expression analysis is crucial. Moreover, the approach of gene silencing can be used if the most differentially expressed genes are known. By “silencing” or inhibiting particular genes linked to particular diseases, gene silencing is a relatively new therapy method that makes use of the body's natural processes to control the disease. Nevertheless, due to their high dimensions, complexity, and presence of duplicate feature values, studies of RNA gene expression data are highly difficult [11]. As a result, automatic feature extraction is required, and machine learning (ML) and deep learning algorithms can help with this [12].

Machine learning is a subfield of artificial intelligence that uses learning and experience to detect underlying patterns in data to establish relationships between them [13]. Given that it is getting more and more challenging for humans to identify trends and patterns in data to make predictions about the future. Machine learning (ML) is becoming essential in the age of mass data. Therefore, machine learning is taking the place of humans in the process of identifying underlying patterns in data and forecasting the future to make informed decisions. With essentially no human input, machine learning (ML) retrieves the features it needs to generate predictions. Nearly everywhere is implementing machine learning. Natural language processing, prediction, aviation governance, and biology to identify protein and RNA sequences are some of their common applications [14, 15]. Tumor types and stages can both be classified using machine learning. The remainder of this essay is structured as follows: section two presented the related work. The methodology is illustrated in section three. The experimental results are discussed in section four. Finally, the conclusion presented in section five.

2 Related Work

The promise of ML models employing NGS data for cancer classification has been demonstrated over the last few years. Literature specifically has several works that classify different types of lung cancer based on gene expression data. Joe W. et al. [28] presented a study for Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. The author used five algorithms; PCA, mRMR, DGE, Lasso, and XGboost to select genes. LUAD and LUSC Classification Statistics achieved good accuracies.

2.1 Machine Learning

Since LUAD is the most common type of lung cancer, numerous publications on LUAD and its control classification have been made. A deep learning model employing gene expression from coding RNA and non-coding RNA was presented by Smolander et al. [16]. They were able to achieve a classification accuracy of 94.97% by using RNA coding. Similarly, to this, Fan et al. achieved 91% accuracy using Support Vector Machines (SVMs) and a signature of 12 genes [17]. For the classification, Zhao et al. merged data from ncRNAs, miRNAs, and mRNAs using SVMs. They ultimately settled on 44 genes, and their classification accuracy was 95.3% [18]. Gonzales et al. investigated the differentially expressed genes (DEGs) between SCLC, LUAD, LUSC, and large cell lung carcinoma for the classification of lung cancer subtypes. The performance of several feature selectors and predictive models for categorization were then compared [19]. With the help of k-NN and the Random Forest feature selector, the authors achieved an accuracy of 88.23%.

Wu J et al. [27] presented an analysis of RNA-Sequence data from 110 triple negative and 992 non-triple negative breasts cancer tumor samples from The Cancer Genome Atlas to select 5502 signature genes used in the development and validation of the classification models. They evaluated four different classification models including Support Vector Machines, K-nearest neighbor, Naïve Bayes, and Decision tree using features selected at different threshold levels to train the models for classifying the two types of breast cancer. The classification accuracy was 85%, 87%, 85%, and 90% for triple and non-triple negative breast cancer using K-nearest neighbor (kNN), Naïve Bayes (NGB), Decision trees (DT), and Support Vector Machines (SVM).

2.2 Deep Learning

In recent years, deep learning models have been used in image classification [29, 30], face recognition, and natural language processing. Medical images were classified using deep learning to determine the covid-19 chest or non-covid19 chest [31] covid. Also, the transfer learning of deep learning is used to classify LUAD, LUSC, and control genes as well as forecast mutations, with an area under the curve (AUC) score of 0.978 [20]. With a score of 0.988 AUC for a binary classification problem, Kanavati et al. [21] demonstrated a deep learning model for lung cancer classification utilizing transfer learning. Images that had been expertly labeled were used by authors. A deep learning

model trained on image tiles and subsequent summary statistics extracted from them was used by Graham et al. [22] to offer a two-step process for LUAD, LUSC, and normal classification, achieving an accuracy value of 81%. The classification of image tiles as malignant vs. non-malignant used numerous Convolutional Neural Networks (CNNs) with various architectures, and Li Z. et al. [23] trained and compared their performance, achieving an AUC score of 0.9119.

Schmauch et al. [24] proposed a model called HE2RNA that can be trained to automatically predict RNA-Seq profiles from whole-slide photos by itself, without professional annotation. This model was constructed through the integration of several data types. The study demonstrated an example of how this method might be used for clinical diagnoses, such as the detection of cancers with microsatellite instability. In their study, Rukhsar L. et al. [25] described how normalization and zero padding were employed to transform RNA-Seq results into 2D pictures. Subsequently, using Deep Learning, the pertinent features were extracted and chosen (DL). There must be classified, and eight DL methods were applied. Results and analysis were based on k-fold cross-validation for each DL classifier and four distinct splitting procedures. The study with the greatest overall outcomes used a convolutional neural network. Therefore, of the eight DL models that were examined, CNN performed the best for classification. Francisco Carrillo-Perez et al. [26]. Used histology and RNA-Seq data for lung cancer, squamous-cell cancer, and healthy lung tissue to produce a late fusion categorization model. In comparison to using each source of data separately, the classification model produced results with a mean F1-score of 95.19% and a mean AUC of 0.99 that were superior. When compared to the isolation histology classifier and the isolation gene expression classifier, it can reduce diagnosis error rates by up to 64% and 24%, respectively.

3 Proposed Methodology

3.1 Generate the Gene Expression Matrix

In this work, we used 598 cases. Each sample has a file containing a file ID and some biological information. All cases have their file IDs, probe IDs, transcripts per million (TPM) reads, and gene expression read count. TPM is a value that represents a relative expression level that, in principle, should be comparable between samples [34]. A clinical sheet for samples is attached to the data. Clinical sheet contains the file ID for all cases and their cancer stage; N0, N1, N2, N3, and NX. The first step is matched the probe IDs with gene symbols by using the Ensemble database. The sample file ID is then matched with the case ID in the clinical phenotype file. Samples without clinical phenotyping or labels are excluded from the data sets and are not included in downstream analysis.

The study has used principal component analysis to see the distribution of the samples according to the five cancer stages N0, N1, N2, N3, and NX that exist in the clinical phenotyping file. Figure 1 presents the distribution of samples according to PC1 and PC2; the output of PCA algorithm.

The second step is removing the Duplicated genes by aggregation. At the end of this part, the algorithm outputs the gene expression matrix without any duplicated genes or duplicated clinical phenotyping samples.

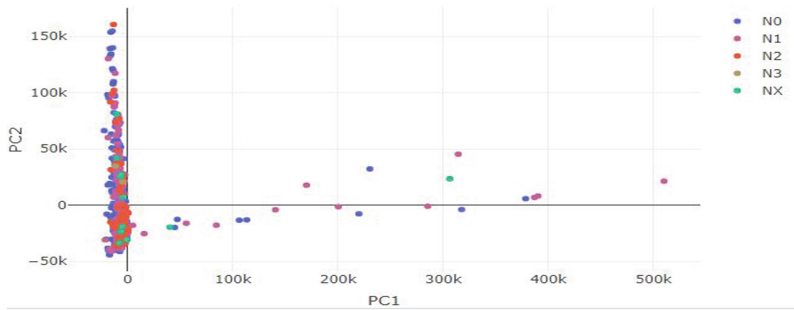


Fig. 1. Principal component analysis for LUAD dataset.

3.2 Differential Gene Expression Analysis and Feature Selection

The gene expression matrix is the input for this stage. All genes’ variances are calculated, and the genes with the lowest variances are eliminated. All genes with a variance less than 3 are eliminated, leaving only the 18308 genes.

The DSeq2 Bioconductor package is used to detect differentially expressed genes, and all the scripts are implemented in R [35]. DSeq2 calculates the log2 fold change (log2 FC), P-values, and p-adjusted values (Padj). Log2-fold change (log2-FC), defined as the median of genes expressed minus the gene expression value for each gene. PADDJ and Log2FC are used to rank genes. Also, these features are selected at different threshold levels. Significantly, we have used two different thresholds for feature selection: the first is “ $\text{padj} < 0.05$ ” and “ $\text{Log2 FC} > \text{Log2}(2)$,” and the second is sorting the top generated genes.” The two-threshold outputs are 12 genes and 5665 genes, respectively. Significantly differentially expressed genes are used to distinguish the five stages of lung cancer. Figure 2 shows a heatmap for 598 samples with 12 genes as the output of the first threshold. The heatmap represents by colors scale the up regulated genes and down regulated genes according to their expression in normal and cancer samples.

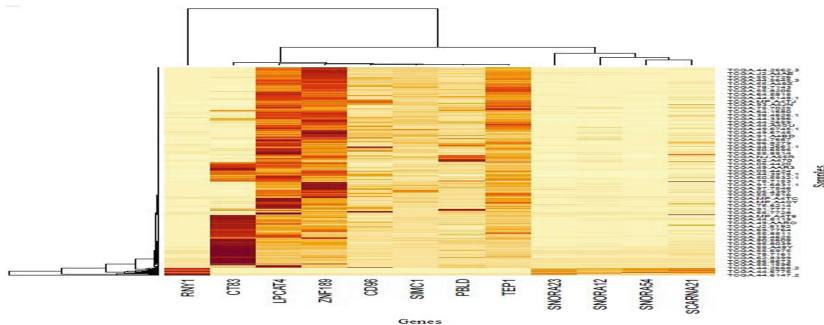


Fig. 2. Heatmap for 598 Samples with 12 DEGs.

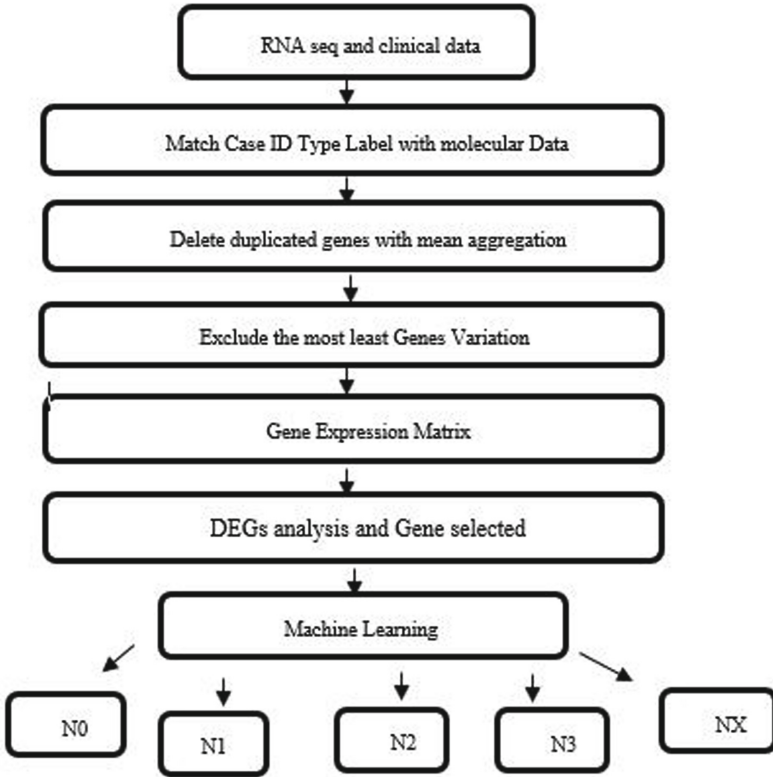


Fig. 3. Block diagram for project design, data processing, analysis workflow, and classification.

3.3 LUAD Stages Classification using Machine Learning

The cancer stage describes the severity of your condition, including the size of the tumor. If it has progressed, the doctor will benefit from knowing the stage of your cancer. In this study, the cancer stage is classified using machine learning algorithms. The machine learning models used in this study are linear regression, logistic regression, RANSAC regression, and decision trees. Each model is applied to 12 genes and 5665 genes to validate their performance. A block diagram for data processes, analysis, and the execution strategy used in this study is presented in Fig. 3.

4 Experimental Results

4.1 Data Set

The used dataset is biological RNA-Seq data. The data are gathered from the Cancer Genome Atlas (TCGA) program [32], which is in the Genomics Data Commons (GDC) portal [33]. Gene expression data and clinical information are obtained from the GDC by using data transfer tool. We have used LUAD data with 598 cases. The data category

is transcriptome profiling. The data type is Gene Expression Quantification, the strategy is RNA-Seq, and the workflow type is STAR-Counts. The data has 334 normal samples and 264 cancer samples. Cancer samples are split into four cancer stages; 98, 75, 41, and 50 samples for N1, N2, N3, and NX Lung cancer stages respectively.

4.2 Results

The study has conducted four different experiments. Each experiment has split the 598 patient samples of LUAD data into 70% training and 30% testing. To validate the gene list, the first experiment has used linear regression algorithm on 598 samples with 12 genes, achieved 100% accuracy. The second experiment has applied a decision tree to 598 LUAD samples for 12 DEGS and achieved 100% test accuracy. We have observed from experiments one and two that linear regression and decision trees validate the 12 DEGs’ with high performance using LUAD cancer stages.

Table 1. Cancer classification comparison between proposed models and the stat-of-the-arts models

Method	Measuring metric	Year
Wu J et al. [27]	Accuracy	2021
K-nearest neighbor (kNN)	0.87	
Naïve Bayes (NGB) 85%	0.85	
Decision trees (DT) 87%	0.87	
Gonzalez S et al. [19]	Relief (Rel) and Random Forest (RF)	2019
SVM	0.85	
k-NN	0.88	
Chen et al. [28]	Accuracy	2021
DGE (Top 500)	0.93	
PCA (Top 500)	0.94	
mRMR (Top 500)	0.91	
Lasso (68 Genes)	0.93	
Xgboost (148 Genes)	0.93	
Proposed Models	12 DEGs (Accuracy)	2022
Linear Regression	1.00	
Logistic Regression	0.91	
Decision Tree	1.00	
RANSAC Regression	0.98	
	5665 DEGs (Accuracy)	
Logistic Regression	1.00	
RANSAC Regression	0.98	

The third experiment has applied logistic regression to the 598 samples for 12 DEGs of gene and 5665 DEGs of gene. The experiment has achieved 90% and 100% for 212 sDEGs-55 and 5665 DEGs, respectively. The fourth experiment RANSAC regression is used on 598 samples in the final experiment for 12 DEGs and 5665 gene DEGs. The experiment achieved 98% for both DEGs. We observed from experiments three and four that logical regression and RANSAC regression validated that the 12 and 5665 genes have high performance in LUAD cancer stages.

Finally, the proposed models are compared to the state-of-the-art: Wu J. et al. [27], Gonzalez S. et al. [19], and Chen et al. [28]. Table 1 shows the comparison between the proposed models and the state-of-the-art models. We have observed that the proposed models validate the 12 and 5665 DEGs of LUAD cancer stages with high performance. Furthermore, when compared to the state-of-the-art, the chosen machine learning algorithms have achieved the highest accuracy.

5 Conclusion

In recent days, cancer has spread greatly, and biologists try to detect the reason for cancer. Many research papers have argued that the main reason for cancer is genes. As a result, we try to use modern science and artificial intelligence to help detect the reasons for cancer. We obtained LUAD data from the GDC portal and then some analyses are done to find the genes with the most variation. DSeq2 is used to calculate Log2 FC and Padj values for genes. Two different thresholds are used to select the most variant genes: 12 DEGs and 5665 DEGs. The study has used different machine learning algorithms to validate the DEGs' performance in classifying the LUAD cancer stages.

The study applied four experiments: the first and second experiments applied linear regression and decision trees to 12 DEGs, achieving 100% accuracy for the two models. The third and fourth experiments applied logistic regression and RANSAC regression to 12 DEG and 5665 DEG, respectively. The two experiments achieved 91% and 98% for 12 DEGS and 100% and 98% for 5665 DEGS for the two models, respectively. Finally, we have compared the proposed methodology and machine learning models with the state-of-the-art and conclude that the analysis has succeeded in identifying the most effective genes in LUAD cancers. Furthermore, when compared to the state-of-the-art, the selected machine learning models validate the performance of the output DEGs with the highest accuracy.

References

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., et al.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* **71**, 209–249 (2021)
2. Xiao, Y., Wu, J., Lin, Z., Zhao, X.: A semi-supervised deep learning method based on stacked sparse auto-encoder for cancer prediction using RNA-seq data. *Comput. Methods Programs Biomed.* **166**, 99–105 (2018)
3. Organization WH. World Cancer Report (2014)
4. UK CR. Types of lung cancer. <https://www.cancerresearchuk.org/about-cancer/lung-cancer/stages-typesgrades/types>

5. Subramanian, J., Govindan, R.: Lung cancer in never smokers: a review. *J Clin Oncol.* **25**(5), 561–570 (2007)
6. Zararsiz, G., Goksuluk, D., Korkmaz, S., Eldem, V., Goksuluk, I.P., Unver, T.: MLSeq machine learning interface to RNA-seq data. <https://bioconductor.org/packages/release/bioc/vignettes/MLSeq/inst/doc/MLSeq.pdf> Accessed 1 July 2021
7. Waseem, Q., Alshamrani, S., Nisar, K., Din, W.W., Alghamdi, A.: “Future technology”, software-defined network (SDN) forensic. *Symmetry* **13**, 767 (2021)
8. Wesolowski, S., Birtwistle, M.R., Rempala, G.A.: A comparison of methods for RNA-seq differential expression analysis and a new empirical bayes approach. *Biosensors* **3**, 238–258 (2013)
9. Conesa, A., et al.: A survey of best practices for RNA-seq data analysis. *Genome Biol* **17**, 13 (2016)
10. Urda, D., Montes-Torres, J., Moreno, F., Franco, L., Jerez, J.M.: Deep learning to analyze RNA-seq gene expression data. In: Rojas, I., Joya, G., Catala, A. (eds.) IWANN 2017. LNCS, vol. 10306, pp. 50–59. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59147-6_5
11. Danaee, P., Ghaeini, R., Hendrix, D.A.: A deep learning approach for cancer detection and relevant gene identification. *Pac. Symp. Biocomput.* **22**, 219–229 (2017)
12. Khattak, A., Asghar, M.Z., Ishaq, Z., Bangyal, W.H., Hameed, A.I.: Enhanced concept-level sentiment analysis system with expanded ontological relations for efficient classification of user reviews. *Egypt. Inform. J.* **22**, 455–471 (2021)
13. Angra, S., Ahuja, S.: Machine learning and its applications a review. In: International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, India, pp. 57–60 (2017)
14. Langley, P., Simon, H.A.: Applications of machine learning and rule induction. *Commun. ACM* **38**, 54–64 (1995)
15. Memon, S.K., Nisar, K., Hijazi, M.H.A., Chowdhry, B.S., Sodhro, A.H.: A survey on 802.11 mac industrial standards, architecture, security & supporting emergency traffic: future directions. *J. Ind. Inf. Integr.* **24**, 100225 (2021)
16. Smolander, J., Stupnikov, A., Glazko, G., Dehmer, M., Emmert-Streib, F.: Comparing biological information contained in mRNA and non-coding RNAs for classification of lung cancer patients. *BMC Cancer* **19**(1), 1176 (2019)
17. Fan, Z., Xue, W., Li, L., Zhang, C., Lu, J., Zhai, Y., et al.: Identification of an early diagnostic biomarker of lung adenocarcinoma based on co-expression similarity and construction of a diagnostic model. *J. Transl. Med.* **16**(1), 205 (2018)
18. Zhao, J., Cheng, W., He, X., Liu, Y., Li, J., Sun, J., et al.: Construction of a specific SVM classifier and identification of molecular markers for lung adenocarcinoma based on lncRNA-miRNA-mRNA network. *Onco Targets Ther.* **11**, 3129 (2018)
19. González, S., Castillo, D., Galvez, J.M., Rojas, I., Herrera, L.J.: Feature selection and assessment of lung cancer sub-types by applying predictive models. In: Rojas, I., Joya, G., Catala, A. (eds.) IWANN 2019. LNCS, vol. 11507, pp. 883–894. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20518-8_73
20. Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyo, D., et al.: Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**(10), 1559–1567 (2018)
21. Kanavati, F., Toyokawa, G., Momosaki, S., Rambeau, M., Kozuma, Y., Shoji, F., et al.: Weakly-supervised learning for lung carcinoma classification using deep learning. *Sci. Rep.* **10**(1), 1–11 (2020)
22. Graham, S., Shaban, M., Qaiser, T., Koohbanani, N.A., Khurram, S.A., Rajpoot N.: Classification of lung cancer histology images using patch-level summary statistics. In: Medical Imaging: Digital Pathology, vol. 10581, p. 1058119 (2018). International Society for Optics and Photonics

23. Li, Z., Hu, Z., Xu, J., Tan, T., Chen, H., Duan, Z., et al.: Computer-aided diagnosis of lung carcinoma using deep learning-a pilot study. arXiv preprint arXiv: 18030 5471 (2018)
24. Schmauch, B., Romagnoni, A., Pronier, E., et al.: A deep learning model to predict RNA-seq expression of tumours from whole slide images. *Nat. Commun.* **11**, 3877 (2020). <https://doi.org/10.1038/s41467-020-17678-4>
25. Rukhsar, L., Bangyal, W.H., Ali Khan, M.S., Ag Ibrahim, A.A., Nisar, K., Rawat, D.B.: Analyzing RNA-seq gene expression data using deep learning approaches for cancer classification. *Appl. Sci.* **12**, 1850 (2022)
26. Carrillo-Perez, F., et al.: Non-small-cell lung cancer classification via RNA-Seq and histology imaging probability fusion. *BMC Bioinform.* **22**, 1–19 (2021)
27. Wu, J., Hicks, C.: Breast cancer type classification using machine learning. *J. Personalized Med.* **11**(2), 61 (2021). <https://doi.org/10.3390/jpm11020061>
28. Chen, J.W., Dhahbi, J.: Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. *Sci. Rep.* **11**, 13323 (2021). <https://doi.org/10.1038/s41598-021-92725-8>
29. Ghaleb, M.S., Ebied, H.M., Shedeed, H.A., Tolba, M.F.: Image retrieval based on self-organizing feature map and multilayer perceptron neural networks classifier. In: 9th International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, pp. 189–193 (2019)
30. Ghaleb, M.S., Ebied, H.M., Shedeed, H.A., Tolba, M.F.: Image retrieval based on deep learning. *J. Syst. Manag. Sci.* **12** (2022)
31. Ghaleb, M.S., Ebied, H.M., Shedeed, H.A., Tolba, M.F.: COVID-19 X-rays model detection using convolution neural network. In: Artificial Intelligence and Computer Vision, AICV 2021, vol. 1377. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_1
32. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., et al.: The cancer genome atlas pancancer analysis project. *Nat. Genet.* **45**(10), 1113 (2013)
33. Grossman, R.L., Heath, A.P., Ferretti, V., Varmus, H.E., Lowy, D.R., Kibbe, W.A., et al.: Toward a shared vision for cancer genomic data. *New Engl. J. Med.* **375**(12), 1109–12 (2016)
34. Li, B., Dewey, C.N.: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011). <https://doi.org/10.1186/1471-2105-12-323>
35. Ritchie, M.E., Phipson, B., Wu, D.: Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015)



Deep Learning for Image and Sound Data: An Overview

Hilali Manal¹(✉), Ezzati Abdellah¹, and Ben Alla Said²

¹ Faculty of Sciences and Techniques, Hassan First University, Settat, Morocco
m.hilali@uhp.ac.ma

² National School for Applied Science, Hassan First University, Berrchid, Morocco

Abstract. The growth of deep learning (DL) in recent years has been astonishing, and every domain in our life is often touched and sometimes consumed with DL breakthroughs. This brought us to why this state-of-the-art was elaborated, the algorithms used in DL are reviewed, their different architectures and designs, and the application domains are based on the type of input data, namely image and sound data. The domains of image classification and object detection for the former, and the domain of automatic speech recognition, Automatic spoken language identification, and Speech emotion recognition for the latter. Hoping through this paper that the task of looking for the latest techniques in DL is made easier for new researchers.

Keywords: Deep learning (DL) · Image data · Sound data · Object detection

1 Introduction

Deep learning methods have proved much superior performance in many traditional computer vision applications such as object detection, classification, and segmentation [1–4]. It is used in driverless car testing, stores, medical diagnostics, and monitoring crops' health. It also succeeded greatly in other domains, such as remote sensing [5], robotics, virtual Assistants, and fraud Detection.

The use of DL models is, day after day, getting bigger. The number of contributions published in the domains is huge, which makes keeping up with the recent work a challenge for many researchers, thus the contribution of this article trying to make looking for the latest techniques in DL based on input data, namely image, and sound data, easier for the new researchers.

This paper is organized as follows: in Sect. 2, a brief presentation of related work. Then in Sect. 3, the most common deep neural network algorithms used in all the fields are overviewed. Application areas models for two types of data: image data, namely image classification, object detection, and sound data are discussed in Sect. 4. Finally, Sect. 5 concludes, and new horizons are presented through future work.

2 Related Work

Many DL reviews have been published on object detection [1], where the author tries to provide a comprehensive survey on the detection framework, the training strategies, and some other aspects of object detection. On image classification [2], the focus was on the efficiency of CNN in the field. On image and semantic video segmentation [3, 4], a review of the challenges of the field, the existent methods and the datasets, and remote sensing [5].

Even though there are many excellent literature reviews on each sub-subject, the articles that give a global idea about this subject are lacking. Thus, this work is believed to be original in presenting recent DL breakthroughs based on input data, namely image and sound data.

3 Deep Learning Advancement Based on Algorithms

It is imperative to introduce, at first, the concepts that back up this overview to make the comprehension of this overview easier for the reader. And to list the methods, designs, and known model networks used as the foundation for deep learning.

Since only the algorithms used as the base for the models in the coming sections will be explained, many other algorithms aren't cited in this overview, such as autoencoders, stacked autoencoders, Restricted Boltzmann machines [6], and deep belief networks.

3.1 Convolutional Neural Networks

CNN, the most popular DL model, is suitable for processing remote-sensing image data. Nonetheless, it is used in other applications that will be elaborated on in the coming sections of this review. The CNN consists mainly of convolution, pooling, and fully connected layers, and the input is processed in each layer to produce a feature map. Some of the popular CNN architectures are AlexNet [7], Visual Geometry Group (VGG) [8] network, Microsoft's ResNet [9], and also GoogLeNet [10].

3.2 Recurrent Neural Networks

The RNN model is used for sequence analysis. This is why the tasks that use sequential data as input, e.g., speech and language processing, use the RNN more. Not only that but most of the backpropagation techniques are used to train RNNs. Storing the information is problematic for this model. To amend this, several memory units have been developed, e.g., the long short-term memory cell (LSTM) [11] and gated recurrent unit (GRU) [12]. They are applied in predicting characters in sequence and even in more complex tasks of remote-sensing images.

3.3 Generative Adversarial Networks

Generative adversarial networks (GANs) [13] gained popularity in DL. This network has a dual system consisting of a generative and discriminative network. The former generates data while aiming to “mislead” the discriminative by outputting examples that seem realistic and similar to the real data. Then the latter try to discriminate between the generated and the genuine data, and the cycle continues until the discriminative network can’t differentiate between the fake and the real data.

4 Application Areas

Deep learning has been useful in many domains and improved many fields, including healthcare [14]. But in this review, the focus is on the application’s domains by the type of input data used in each one, mainly image data and sound data.

4.1 Image Analysis

Image data has been the core of recent studies, and there was a lot of effort and hard work put into object detection, image classification, generation, and feature extraction. The algorithms reviewed in this section are image classification and object detection algorithms; where image classification’s one are the most widely known in this field, and they surpassed their original purpose to many other tasks and datasets, and the second one is algorithms that use the first one as base as shown in Fig. 1.

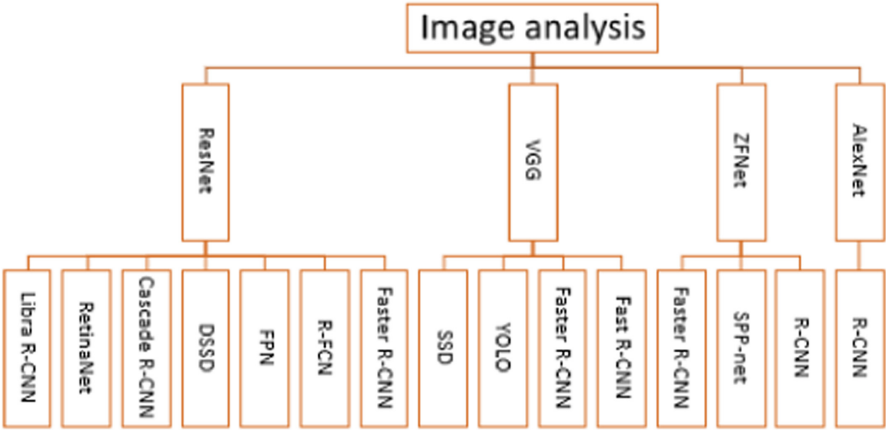


Fig. 1. Image classification and object detection algorithms and the relation between the two

Image Classification. The models existing for this task are numerous. Only the one used as a backbone for the object detection models will be described.

AlexNet [7], the portal to the new era of deep CNN, the model used involves five convolutional layers followed by max-pooling layers using the Rectified Linear Units (ReLU) activation function, then three dense fully-connected layers using ReLU and dropout. It achieved 84.6% accuracy and won the ILSVRC-2012 challenge.

ZFNet [15] altered AlexNet, changing the kernel size of the first layer from 11×11 to 7×7 and the stride size from 4 to 2. Which allowed the model to preserve more low-level features and lessen the downsampling rate. It is well-fitted for large and small object detection and recognition.

A model was introduced by the Visual Geometry Group (VGG) [8] from the University of Oxford. It achieved 92.7% accuracy, which is why it became popular. VGG contains 16 layers which is why it is recognized as VGG-16. It differs from its antecedents in using multiple layers of the convolution layers with few neurons in the first layers as a replacement for a few layers with multiple neurons. This means that the model will contain fewer parameters and more non-linearities in between, which leads to an easier model to train and a more discriminative function.

Microsoft's ResNet [9] is outstanding for winning the ILSVRC-2016 challenge. It achieved 96.4% accuracy. This network is famous for its huge depth (152 layers), and it brought the new idea of residual blocks. To deal with this depth through training, those blocks introduce the "identity skip" networks, allowing the layers to duplicate their inputs to the next layer. It guarantees that the next layer has the unchanged input and the previous layer's output, which gives a new encoded output to the next layer as input. Besides, those connections help to overcome the vanishing gradients problem.

Object Detection. One of the main tasks in image analysis is object detection, aiming to detect targets in an image or video. The contributions in this area can be divided into two categories: the two-stage detectors and the single-stage detectors.

Two-Stage Detectors. The two-stage detector (2sd) detects the object location using a region proposal network (RPN) to define the region of interest (RoI) and then extract the classification of those regions and the location of the target from them.

In region-based CNN(R-CNN) [16], DCNNs are used to extract image features and to select region proposals selection search algorithms (SSA) are used; for the classification of those regions, they used support vector machine (SVM), and finally, for the refinement bounding box was used.

For the extraction of the image features and the selection of region proposal, SPP-net [17] works the same as R-CNN, but it maps the regions to the feature maps, then uses spatial pyramid pooling. Its output is then fed as input to DCNN, this model is $24\text{--}102 \times$ faster than the R-CNN. Another study [18] uses this algorithm as a part detector, then uses a particle swarm optimization approach to examine new regions in the neighborhood of the boxes predicted by SPP-net, and that is to maximize the recognition performance. This study doesn't require a bounding box during testing, but it uses the tow algorithm as a joint part detection and classification task in an iterative manner. In tracking in video analysis, SPP-net can be used to get from each frame the features of the region tracked [19], and combine it with a recurrent R-CNN to capture the information of each region.

Similarly to SPP-net, Fast R-CNN [20] uses DCNNs, an SSA for the same reasons, and generates feature maps, but for the downsampling, it uses an ROI Pooling layer to get fixed-size feature maps, it is $3\text{--}10 \times$ faster than SPP-net.

Faster R-CNN [21] uses RPN to replace the SSA, this network shares the feature maps with the base network. Combined with the ResNet-50 network, this algorithm was used to get a better score detecting two maturity stages of coconuts on real-time images [22].

R-FCN [23] construct a set of position-sensitive score maps, then on top of an FCN, append a position-sensitive RoI pooling layer that shepherds information from these score maps. The entire architecture is learned end-to-end.

FPN [24] is a proposed multi-level feature fusion Feature Pyramid Network that is beneficial to multi-scale and small object detection.

Cascade R-CNN [25], another version of R-CNN with multiple stages, uses the output of one stage to train the next. This algorithm is used to refine the bounding boxes iteratively combined with FPN as a feature extractor [26] for detecting objects in different lighting conditions.

Among the models cited, the most used ones are the R-CNN and its family of Fast and Faster R-CNN, whereas FPN is mostly used combined with other algorithms or its capability in small object detection.

Single-Stage Detectors(1sd). They are the algorithms that combine the object location and classification in a single stage, and they directly detect the objects from the anchors instead of using RPN.

YOLO [27] proposes a new single-stage detection network. It is fast and can meet real-time requirements. It was combined with FPN since it increased the detection accuracy of small objects when applied to YOLOv2, but detecting large objects was difficult for it. Thus, a better version of YOLO, YOLOv3 used Duplex FPN [28], to amend this problem and detect large objects more accurately.

SSD [29] uses a Multi-layer detection mechanism and a Multi-scale anchors mechanism in different layers. This model was improved for infrared small target detection of Unmanned Aerial Vehicles [30], called Single Shot MultiBox Detector for Small Target (SSD-ST), by dropping low-resolution layers and enhancing high-resolution layers and designing an Adaptive Pipeline Filter (APF) based on the temporal correlation to further improve the detection results.

Like the SSD, the DSSD [31] model is a Multi-layer mechanism but for feature fusion. It uses deconvolution instead of simple linear interpolation for Up-sampling to enhance the detection accuracy of a small object.

RetinaNet [32] proposes a new loss function Focal Loss for dealing with class imbalance during training. This algorithm was combined with a suite of augmentation techniques and transfer learning for detecting leukocytes in IVM data [33].

Libra R-CNN [34] proposes IoU-balanced sampling, balanced feature pyramid, and balanced L1 loss to balance the imbalance issues in the training process.

According to a study that compares the performances of YOLOv3, Faster R-CNN, and SSD in small object detection [35], Faster R-CNN performed the best, then YOLOv3 came close behind. And between the remaining models, RetinaNet is the most used model with aerial and satellite imagery.

Table 1. Summary of literature on object detection with the databases used in each one.

Model	Category	Base algorithm	Dataset
R-CNN [16]	2sd	AlexNet ZFNet	ILSVRC PASCAL VOC
SPP-net [17]	2sd	ZFNet	ImageNet Pascal VOC Caltech101
Fast R-CNN [20]	2sd	VGGNet	
Faster R-CNN [21]	2sd	VGGNet ResNet ZFNet	PASCAL VOC MS COCO ILSVRC
R-FCN [23]	2sd	ResNet	PASCAL VOC
FPN [24]	2sd	ResNet	COCO
YOLO [27]	1sd	VGGNet DarkNet	PASCAL VOC
SSD [29]	1sd	VGGNet	PASCAL VOC COCO ILSVRC
DSSD [31]	1sd	ResNet	PASCAL VOC COCO
Cascade R-CNN [25]	2sd	ResNet	COCO
RetinaNet [32]	1sd	ResNet ResNext	COCO
Libra R-CNN [34]	1sd	ResNet	MSCOCO

Discussion. Presented in the first part of this section (image data analysis) is the standard and most used deep learning algorithms in image classification. However, those architectures are recently used and fused to adapt to other tasks and processes, which is the case for transfer learning or even in sound data analysis. They have become the core architectures in CNN.

In the second part, different literature on the object detection task is presented, as shown in Table 1, the different models are mostly based on the core architectures that were presented in the previous section, they are separated into single-stage detectors and two-stage detectors with the second ones those that use RPN to define the RoI while the first ones don't, the majority of the models use DL models for the feature extraction or the classification tasks since the focus in this study is on DL algorithms, but some fuses those algorithms with traditional algorithms for either the classification task [16] using SVM or the detection task [18] using PSO as a part detector.

Even though the algorithms were separated into 1sd and 2sd, both use commonly known datasets to be able to compare their contributions, which are either improving the accuracy [27, 29, 31], dealing with class imbalance [32, 34] or to focus on some features [24].

4.2 Sound Data Analysis

Similar to image data, sound data have different applications, such as automatic speech recognition (ASR), Automatic spoken language identification (SLID), or speech emotion recognition (SER). In this section, the description of the different approaches used to develop a signal analyzing algorithm are elaborated as shown in Fig. 2.

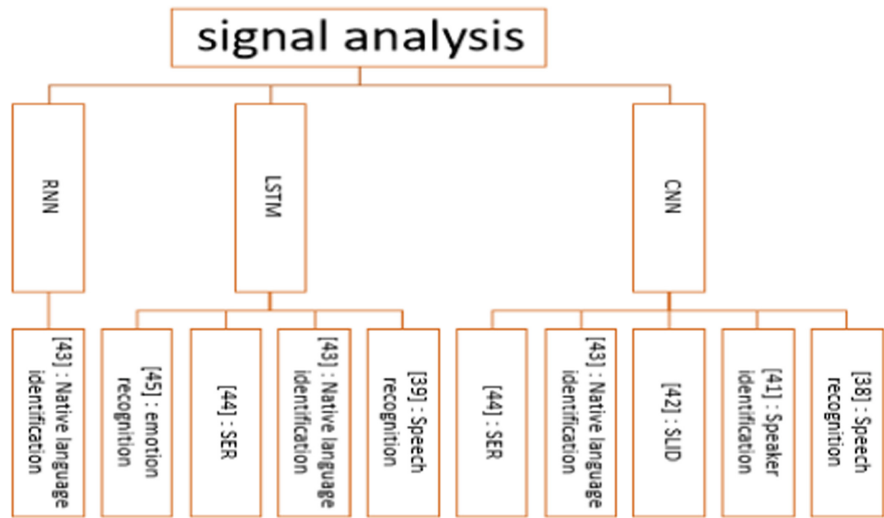


Fig. 2. Signal analysis algorithms, their application areas, and the base architecture behind them

Automatic Speech Recognition. ASR systems, generally split speech recognition tasks. Starting with the extraction of acoustic feature observations from the signal such as (Mel frequency cepstral coefficients (MFCCs) [36] and perceptual linear prediction cepstral features (PLPs) [37] are the features that are typically extracted), this is the first stage. Then, the probability of subwords (generally based on phonemes) is estimated, and based on this estimation and using lexical and syntactical constraints, the best matching word is found. This is the second stage.

The first stage techniques try to purify the feature extracted, such as MFCCs, and get the signals as clean as possible, and it can be used for different speech applications. In the second stage, a few years ago, the Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs) were the state-of-the-art acoustic models (AM) that define the distribution of speech. While recently, AM shifted from GMM to a deep neural network (DNN) due to its capability in representation, the neural network automatically finds the relationship between the noisy speech and the phonetic targets.

An end-to-end model based on CNN [38] was proposed, and the processing of the signal and the extraction of the features is done by multiple convolution layers divided into many filter and classification stages.

An LSTM-based speech recognition model [39] is proposed, and it uses LSTM neural network and CTC [40]. The MFCC features are extracted and then fed to a mini-batch sample selection, and then the arranged samples are fed to an LSTM network that produces probabilities and CTC that decode and calculate the loss.

Voxceleb [41] is a contribution that proposes a deep CNN-based system(VGGVox) for speaker verification, the voice spectrograms are mapped to an embedding space, and the similarity is calculated using the cosine distance. Variants of VGG [8] and ResNet [9] are used as base architecture.

Automatic Spoken Language Identification. There are very few studies in this domain, and two models will be presented, each one different from the other, even though they both identify the spoken language.

FuzzyGCP [42] is a SLID approach, it uses a Deep Dumb Multi-Layer Perceptron as a classifier and a Deep CNN and Semi-Supervised GAN trained on image-based features for maximum to enhance precision and finally applies Ensemble learning to predict the language class.

A study used different end-to-end architectures and attention mechanism variants [43] for automatic native language (French, Arabic, Chinese, German, Hindi, Italian...) recognition from raw features spectrogram of non-native speech (English), namely Bidirectional Gated Recurrent Unit (Bi-GRU) encoder network, LSTM, and CGDNN system that contain a combination of CNN, GRU, and fully connected DNN layers.

Speech Emotion Recognition. Speech emotion recognition isn't a new field of study but due to the recent advancements in technology and Deep Learning it has gained attention, these are some literature research on this topic.

An emotional-category-based feature weighting (ECFW) [44] method Outputs the optimal combination of feature and model by matching each feature and model. It uses the optimal feature selection algorithm to reduce the feature dimension, and it recognizes emotions by calculating the weights for a better feature representation.

A fusion model of LSTM and graph convolutional neural network (GCNN) for emotion classification, a novel deep learning model (ERDL) [45] is proposed, it constructs a feature cube, then this feature cube of each segment serves as input to the model which fuses GCNN and long-short term memories neural networks (LSTM), where multiple GCNNs are applied to extract graph domain features, and the LSTM cells are used to store the change between two channels and extract temporal features, finally, the Dense layer is used for the emotion classification results.

Discussion. In this section, there is the presentation of different literature on sound data analysis, as shown in Table 2, although Deep Learning is mostly used in image data analysis (computer vision, object detection...) or natural language processing, it can handle other tasks such as ASR, SLID, and SER. And in the field of sound data analysis, Deep learning models have made big improvements in speech recognition, multiple systems are easily accessible even through mobile devices, and in contrast, SER systems are yet to be used in real-life scenarios.

Table 2. Summary of the literature on speech analysis with the datasets used and the Deep Learning based algorithms.

Model	Category	Base algorithm	Dataset
CNN-based approach [38]	Speech recognition	CNN-1H (with 1 hidden layer) CNN-3H (with 2 hidden layers)	TIMIT Wall street journal Mediaparl
LSTM-based approach [39]	Speech recognition	LSTM (5 layers 600 units)	VCTK
Voxceleb [41]	Speaker identification	CNN (a modified VGG and ResNet)	VoxCeleb1 VoxCeleb2
FuzzyGCP [42]	SLID	GAN CNN	IIIT Hyderabad (Indian language) IIT Madras (Indian language) VoxForge MaSS
End-to-end architectures for language identification [43]	Native language identification	RNN (LSTM and GRU) CNN	A dataset of non-native English speech collected during a global assessment of English language proficiency
(ECFW) [44]	SER	LSTM CNN	IEMOCAP MASC Berlin (EMO-DB)
(ERDL) [45]	Electroencephalography (EEG) based emotion recognition	Graph CNN LSTM	DEAP

Focusing on the algorithms of Deep learning used in each field and essentially on end-to-end models, it is quite clear that the most used algorithm is CNN [38, 41–45] but it is combined generally with a more suitable sequence model.

The first stage in sound data analysis is feature extraction, and then the second is classification. The features that each model extract differs, but those frequently extracted, other than the traditional MFCCs and PLPs, are voice quality, WPCC, and some features that humans can detect and extract, such as intonation and rhythm.

The datasets that are used differ from one model to another and are multiple. Some use known models to compare their results [44–46], while others prefer to create their datasets to adapt to the specifications of their studies [43, 47] or amend the lack in some public datasets [41].

5 Conclusion and Future Direction

This study presents an effort to summarize different deep learning models, starting with the advancement made in this field based on algorithms, which are crucial since they are the pillars of each model presented in the application area part.

Then followed by defining the application area based on the type of data fed to the model; for image data, the talk was about image classification, which is based on CNN, then about object detection with its two types of single-stage and two-stage detectors, listing the most common models and their field of expertise and different studies that used the same algorithm as a backbone or as the main model.

Finally, concluding with sound data in three different fields, namely Automatic speech recognition, Automatic spoken language identification, and emotion recognition, each with some recent algorithms.

Future work would focus on implementing a DL end-to-end model to analyze a signal input using CNN fused with LSTM. Researchers might also consider using the widely known image classification algorithm as a base algorithm, similar to how object detection algorithms cited in Sect. 3 used them.

References

1. Liu, L., et al.: Deep learning for generic object detection: a survey. *Int. J. Comput. Vis.* **128**(2), 261–318 (2019). <https://doi.org/10.1007/s11263-019-01247-4>
2. Sornam, M., Muthusubash, K., Vanitha, V.: A survey on image classification and activity recognition using deep convolutional neural network architecture. In: 2017 9th International Conference on Advanced Computing (ICoAC), pp. 121–126 (2017). <https://doi.org/10.1109/ICoAC.2017.8441512>
3. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J.: A review on deep learning techniques applied to semantic segmentation. [arXiv:1704.06857](https://arxiv.org/abs/1704.06857) [cs] (2017)
4. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., Garcia-Rodriguez, J.: A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput.* **70**, 41–65 (2018). <https://doi.org/10.1016/j.asoc.2018.05.018>
5. Yuan, X., Shi, J., Gu, L.: A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* **169**, 114417 (2021). <https://doi.org/10.1016/j.eswa.2020.114417>
6. Nekrasov, P., Freeze, J., Batista, V.: Using restricted boltzmann machines to model molecular geometries. [arXiv:2012.06984](https://arxiv.org/abs/2012.06984) [physics] (2020)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105. Curran Associates, Inc. (2012)
8. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) [cs] (2015)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. [arXiv:1512.03385](https://arxiv.org/abs/1512.03385) [cs] (2015)

10. Szegedy, C., et al.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9. IEEE, Boston, MA, USA (2015). <https://doi.org/10.1109/CVPR.2015.7298594>
11. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
12. Cho, K., et al.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. [arXiv:1406.1078](https://arxiv.org/abs/1406.1078) [cs, stat] (2014)
13. Goodfellow, I.J., et al.: Generative adversarial networks. [arXiv:1406.2661](https://arxiv.org/abs/1406.2661) [cs, stat] (2014)
14. Rahman, S., Wang, L., Sun, C., Zhou, L.: Deep learning based HEp-2 image classification: A comprehensive review. *Med. Image Anal.* **65**, 101764 (2020). <https://doi.org/10.1016/j.media.2020.101764>
15. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. [arXiv:1311.2901](https://arxiv.org/abs/1311.2901) [cs] (2013)
16. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. [arXiv:1311.2524](https://arxiv.org/abs/1311.2524) [cs] (2014)
17. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8691, pp. 346–361. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10578-9_23
18. Huang, C., He, Z., Cao, G., Cao, W.: Task-driven progressive part localization for fine-grained object recognition. *IEEE Trans. Multimedia* **18**, 2372–2383 (2016). <https://doi.org/10.1109/TMM.2016.2602060>
19. Kong, L., Huang, D., Qin, J., Wang, Y.: A joint framework for athlete tracking and action recognition in sports videos. *IEEE Trans. Circuits Syst. Video Technol.* **30**, 532–548 (2020). <https://doi.org/10.1109/TCSVT.2019.2893318>
20. Girshick, R.: Fast R-CNN. [arXiv:1504.08083](https://arxiv.org/abs/1504.08083) [cs] (2015)
21. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. [arXiv:1506.01497](https://arxiv.org/abs/1506.01497) [cs] (2016)
22. Parvathi, S., Selvi, S.T.: Detection of maturity stages of coconuts in complex background using Faster R-CNN model. *Biosyst. Eng.* **202**, 119–132 (2021). <https://doi.org/10.1016/j.biosystemseng.2020.12.002>
23. Dai, J., Li, Y., He, K., Sun, J.: R-FCN: object detection via region-based fully convolutional networks vol. 9
24. Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. [arXiv:1612.03144](https://arxiv.org/abs/1612.03144) [cs] (2017)
25. Cai, Z., Vasconcelos, N.: Cascade R-CNN: delving into high quality object detection. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6154–6162. IEEE, Salt Lake City, UT (2018). <https://doi.org/10.1109/CVPR.2018.00644>
26. Wu, C.-E., Chan, Y.-M., Chen, C.-H., Chen, W.-C., Chen, C.-S.: IMMVP: an efficient day-time and nighttime on-road object detector. In: 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP 2019). IEEE, New York (2019)
27. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. [arXiv:1506.02640](https://arxiv.org/abs/1506.02640) [cs] (2016)
28. Shin, S., Han, H., Lee, S.H.: Improved YOLOv3 with duplex FPN for object detection based on deep learning. *Int. J. Elec. Eng. Educ.* <https://doi.org/10.1177/0020720920983524>
29. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
30. Ding, L., Xu, X., Cao, Y., Zhai, G., Yang, F., Qian, L.: Detection and tracking of infrared small target by jointly using SSD and pipeline filter. *Digital Signal Process.* **110**, 102949 (2021). <https://doi.org/10.1016/j.dsp.2020.102949>

31. Fu, C.-Y., Liu, W., Ranga, A., Tyagi, A., Berg, A.C.: DSSD: deconvolutional single shot detector. [arXiv:1701.06659](https://arxiv.org/abs/1701.06659) [cs] (2017)
32. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. [arXiv:1708.02002](https://arxiv.org/abs/1708.02002) [cs] (2018)
33. da Silva, B.C.G., Tam, R., Ferrari, R.J.: Detecting cells in intravital video microscopy using a deep convolutional neural network. *Comput. Biol. Med.* **129**, 104133 (2021). <https://doi.org/10.1016/j.combiomed.2020.104133>
34. Pang, J., Chen, K., Shi, J., Feng, H., Ouyang, W., Lin, D.: Libra R-CNN: towards balanced learning for object detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 821–830. IEEE, Long Beach, CA, USA (2019). <https://doi.org/10.1109/CVPR.2019.00091>
35. Liu, Y., Sun, P., Wergeles, N., Shang, Y.: A survey and performance evaluation of deep learning methods for small object detection. *Expert Syst. Appl.* **172**, 114602 (2021). <https://doi.org/10.1016/j.eswa.2021.114602>
36. Logan, B.: Mel frequency cepstral coefficients for music modeling. In: International Symposium on Music Information Retrieval (2000)
37. Honig, F., Stemmer, G., Hacker, C., Brugnara, F.: Revising perceptual linear prediction (PLP), vol. 4 (2005)
38. Palaz, D., Magimai-Doss, M., Collobert, R.: End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition. *Speech Commun.* **108**, 15–32 (2019). <https://doi.org/10.1016/j.specom.2019.01.004>
39. Dokuz, Y., Tufekci, Z.: Mini-batch sample selection strategies for deep learning based speech recognition. *Appl. Acoust.* **171**, 107573 (2021). <https://doi.org/10.1016/j.apacoust.2020.107573>
40. Graves, A., Fernández, S., Gomez, F., Schmidhuber, J.: Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 369–376. Association for Computing Machinery, New York, NY, USA (2006). <https://doi.org/10.1145/1143844.1143891>
41. Nagrani, A., Chung, J.S., Xie, W., Zisserman, A.: Voxceleb: large-scale speaker verification in the wild. *Comput. Speech Lang.* **60**, 101027 (2020). <https://doi.org/10.1016/j.csl.2019.101027>
42. Garain, A., Singh, P.K., Sarkar, R.: FuzzyGCP: a deep learning architecture for automatic spoken language identification from speech signals. *Expert Syst. Appl.* **168**, 114416 (2021). <https://doi.org/10.1016/j.eswa.2020.114416>
43. Ubale, R., Qian, Y., Evanini, K.: Exploring end-to-end attention-based neural networks for native language identification. In: 2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece (2018). <https://doi.org/10.1109/SLT.2018.8639689>
44. Li, D., Zhou, Y., Wang, Z., Gao, D.: Exploiting the potentialities of features for speech emotion recognition. *Inf. Sci.* **548**, 328–343 (2021). <https://doi.org/10.1016/j.ins.2020.09.047>
45. Yin, Y., Zheng, X., Hu, B., Zhang, Y., Cui, X.: EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. *Appl. Soft Comput.* **100**, 106954 (2021). <https://doi.org/10.1016/j.asoc.2020.106954>
46. Huang, Y., Tian, K., Wu, A., Zhang, G.: Feature fusion methods research based on deep belief networks for speech emotion recognition under noise condition. *J. Ambient. Intell. Humaniz. Comput.* **10**(5), 1787–1798 (2017). <https://doi.org/10.1007/s12652-017-0644-8>
47. Huang, H.B., Huang, X.R., Li, R.X., Lim, T.C., Ding, W.P.: Sound quality prediction of vehicle interior noise using deep belief networks. *Appl. Acoust.* **113**, 149–161 (2016). <https://doi.org/10.1016/j.apacoust.2016.06.021>



Correction of an Image for Colour Blindness Using the Fusion of Ishihara Filter and Histogram Equalization

M. S. Sannidhan¹, Jason Elroy Martis², C. V. Aravinda¹, and Roheet Bhatnagar³(✉)

¹ Department of CSE, NMAM Institute of Technology, Nitte 574110, India
{sannidhan, aravinda.cv}@nitte.edu.in

² Department of ISE, NMAM Institute of Technology, Nitte 574110, India
jason1987martis@nitte.edu.in

³ Department of CSE, Manipal University Jaipur, Jaipur 303007, Rajasthan, India
roheet.bhatnagar@jaipur.manipal.edu

Abstract. Colour blindness is one of the minor disorders in individuals appearing due to a genetic mutation causing the intensity to few colours making the individual fail to discriminate specific colours. Concerning the nobility of colour recognition, the disorder ranges in differentiating monochromatic colour to the most common form. This ‘red-green’ dissimilarity causes vibrant colours to override dull colours easily are easily seen, and the dull shades are difficult to perceive. Targeting the impacts of colour blindness, in our research work, we have proposed a designed colour filter built on the combination of the well-known Ishihara test t and histogram equalization. Ishihara filter alone is insufficient as the approach is not suited well to translate real-world images. To overcome this problem, the proposed work aimed at the modification Ishihara filter eradicating lighter shades and shifting them to the darker regions. The image obtained after applying histogram equalization could also be perceived by a colour-blind person. The research paper has also presented a comparative analysis of the two approaches for the elimination of colour blindness.

Keywords: Colour blindness · Image processing · Ishihara test · Colour correction · Ishihara filter · Histogram equalization

1 Introduction

Colour blindness is one of the naturally occurring vision deficiencies to discriminate a few colours based on genetic defects. It is most challenging for people with this deficiency to identify the objects within an image when it has a range of colour combinations. Hence there is an able demand for implementing a system that could aid in overcoming colour deficiency. Colour blindness is also one of the pains in modern society, as people with this deficiency have many problems even in day-to-day activities. In this regard, our proposed system targets correcting any digital image with able filters that overcome the problem of colour overlap for any colour-deficient individual.

A colour blindness deficiency is not life-threatening; people with this kind of abnormality can also lead an everyday life like any other individual. Many of those exposed to exhaustive tests will not even know about their deficiency. However, the modern era has considered this problem a concern to aid colour-blind people, and many areas have even taken specific measures. One among them is displaying the traffic signal light by adding blue hues to the green colour and red with an orange hue. With the invention of modern image processing techniques, it is now even possible to apply the same concept for any digital image to enhance the perception of colour-blind people [1–4].

In this research work, we have focused on the enhancement of the image using two different techniques 1) one being the Ishihara filter and 2) the other using Histogram equalization to correct the colour blindness that is being induced within the digital image. The primary objective of the proposed system is to enhance an image so that any individual can rightly perceive it. Hence, the identified image enhancement techniques are used as a pre-processor for improving the input image quality in the initial step [5–7]. Two standard techniques: 1) Ishihara filter and 2) histogram equalization techniques are implemented to obtain the enhanced image to correct the colour blindness from the image. Since the two techniques implemented for colour blindness correction yield varying results in quality assessment, a comparative analysis of both techniques is carried out to the result produced by them. To validate the accurate result depending on the type of image given as input and for examining some dependent constraints.

2 Related Works

Research article [8, 9] discussed different image enhancement procedures for the characteristic of colour blindness that is biologically related. The work presented the implementation of a colour filter for an image in connection to the Ishihara colour tests. The filter was applied over a digital image, and the work's outcomes proved the proposed system is better working capability. Two different tools were designed in the article [10] to help people overcome colour blindness due to overlaps in colour hues. The first technique uses a test, and then based on the tests, a Fuzzy Logic is designed to simulate red, green and blue colours. Experimental outcomes were limited to simulating only two colours without variation and were computationally expensive. Research article [11] implemented a local minima-based image enhancement to enhance the images in the initial step. The proposed method computes any input image's histogram and provides the result that could analyze in grey levels for precise object detection. However, the image with overlapping contrast is a challenge to this technique due to more noise. Hence the improvement is accomplished in A novel idea article [12]. Researchers in the article [13] developed a colour-sensing system for colour-blind people. The technique is a hardware-based system with a microcontroller that detects and reveals the colour based on voltage values. However, the technique is appropriate in real-time scenarios.

An idea was developed in an article [14] to test colour blindness in individuals. The technique only covered the distinction between primary and secondary colours. A recolouring scheme was discussed in the article [15] that utilized the idea of modifying the colours to suit the ability of a colour-blind individual. Authors have designed an algorithm that develops a model for recording differences in colour values in connection

to colour-blind people. The approach is simple but unsuitable when there is a significant variation in colour values. Another article [16] discussed a technique that combines a median filter and a noise detector. A median filter is used to filter out noisy pixels. The algorithm works well for suppressing impulse noise with noise density ranging from 10 to 70% while preserving image details. However, the technique again fails when there is a significant variation in colour components degrading the quality of the image.

Article [17, 18] is another well-known work that demonstrated the histogram equalization method as a pre-processor for the original image to enhance the image and achieve helpful information. The employment of histogram equalization ensured the clarity of colour components. Work carried out in the article [19] considered enhancing the images via sharpening techniques. Psychometric-based analysis was presented in the outcomes that would generate a score based on observer evaluation. Evaluations proved the technique worked better in enhancing the image but was not suitable for sharp variations in the colour.

3 Proposed System

The proposed system comprises of two important modules: Ishihara filter and histogram equalization. Figure 1 illustrates the overview of the system in its entirety and its modules are explained in detail in the following subsections.

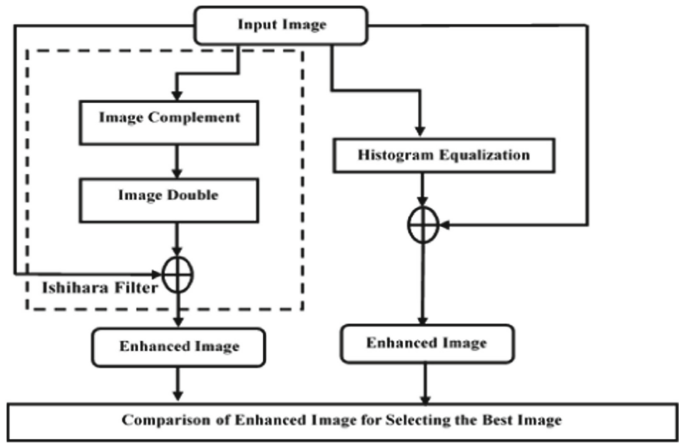


Fig. 1. Overview of the proposed system

3.1 Pre-processing

The pre-processing step's primary goal is to enhance an image that improves the quality of the image's features useful for next-level processing by overcoming unwanted distortions. Hence it is used to improve certain features for subsequent analysis or for image display. Effortlessly, in this step, each pixel of an image is processed to rule out the green and blue components keeping only the red colour components. The step is depicted with input and its resultant in Fig. 2.

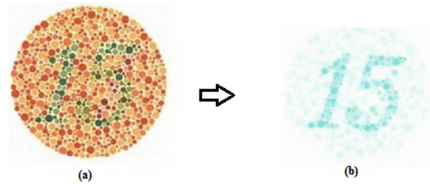


Fig. 2. Preprocessing: (a) original image (b): pre-processed image

3.2 Image Enhancement Techniques

This step concentrates on processing the supplied image so that the resulting image becomes more appropriate than its original form for next-level processing by enhancing an image's boundaries, contrast and edges, making its graphic display more compatible with further processing [19]. Mentioned procedures under the enhancement techniques of the image processing are available as a standard filter. Filtering is used to modify or enhance an image by emphasizing certain core features or removes other features that distort or harm the quality of an image. There are a good number of standard filters in image processing. Based on the exhaustive review and the nature of the problem, it was made evident that a median filter would perform better. Hence the same is chosen. The median filter separately considers each pixel in the image and its immediate neighbors to decide if its surroundings are representative. Further, by numerically sorting the pixel values, the median is computed from the pixels of adjacent neighborhoods, and the median pixel value is used to replace the other pixel being considered. The procedure is mathematically formulated as presented in Eq. (1)

$$f(x, y) = \text{median}(s, t) \in S_{xy}\{g(s, t)\} \quad (1)$$

In Eq. (1), the term $g(s, t)$ is the pixel corresponding to the original image and $f(x, y)$ is the pixel representing the resulting filtered image. The median filter extends two methods in filtering: Image complement and Image Doubling.

• Image Complement

In this procedure, a pre-processed image is taken as input and is processed to obtain its complement. By this, dark regions of the output image appear lighter and light regions appear darker. Output of this stage is negated image. Figure 3 depicts the outcomes of the procedure.

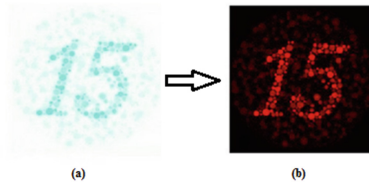


Fig. 3. Image complement (a) pre-processed image (b): negated image

• Doubling an Image

In this stage, an integer class value of each pixel of a negated image is then converted into a double precision value. This doubles the intensity of the pixels for any colour leaving behind the dark region. Hence, the stage delivers an image that appears brighter, ensuring a better detail than the image before processing. Figure 4 shows the enhanced version of the negated image on doubling it.

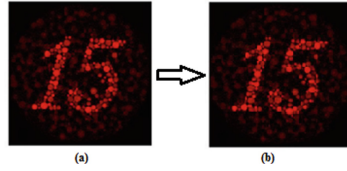


Fig. 4. Image doubling: (a) negated image (b) doubled image

3.3 Histogram Equalization

We have parallelly implemented Histogram equalization method over the original colour-blinded image. The technique aids in enhancing the lower contrast area of an image to a higher level. Technique remaps the intensity of each pixel value corresponding to source image such that processed image's pixels have uniformity in the histogram concerning its intensity. The technique mainly aims to stretch contrast at the darker regions and contrast shrinking at the brighter regions. Therefore, the processed image will be more detailed with an addition of a minimum noise [11]. Mathematically the concept of histogram equalization is formulated in Eq. (2).

$$g(i, j) = \text{floor}((L - 1) \sum_{n=0}^{f(i, j)} P_n) \quad (2)$$

Function ' f ' represents the initial image before processing, g represents the image after applying the procedure, L is the highest pixel intensity which is often considered to be 256, P denotes the normalized histogram for f concerning an intensity value n . Mathematical formulation for the same is presented in Eq. (3)

$$P_n = \frac{\text{number of pixels with intensity value } n}{\text{total number of pixels}} \quad (3)$$

The result of processing performed through the histogram equalization technique is visually presented in Fig. 5.

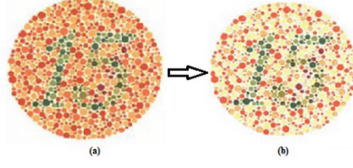


Fig. 5. Histogram equalization technique: (a) original image (b) histogram equalized image

4 Experiments and Results

Experiment is conducted on colour-blinded images of Ishihara plate benchmark data set. Based on the three standard image quality assessment methods, MSE, PSNR and SSIM, we choose a technique that yields a better result for a particular image. An experiment is conducted on every image and their result is presented in this section. Following are the description of quality metrics [20–23].

- **PSNR (Peak Signal-to-Noise Ratio)**

PSNR calculates the highest power of an image concerning the generated noise. Mathematically the quality metric is formulated in Eq. (4)

$$PSNR = 10 \log_{10} \left(\frac{peakval^2}{MSE} \right) \quad (4)$$

In Eq. (4), peakval is Peak Signal Value and MSE is explained in Eq. (5).

- **MSE (Mean Squared Error)**

This evaluator computes average error values. Equation (5) describes the mathematical formulation for the same.

$$MSE = \frac{\sum_{m=0, n=0}^{m=M, n=N} [I_1(m, n) - I_2(m, n)]^2}{M * N} \quad (5)$$

In equation M, N denotes the number of rows and columns of the image.

- **SSIM (Structural Similarity Index)**

This method gauges the quality of an image in digital format. It does so by evaluating the similarities between the two images. The mathematical formulation for the same is given in Eq. (6)

$$SSIM(i, j) = \frac{(2 * \mu_1 * \mu_1 + c_1) * (2 * \sigma_{ij} + c_2)}{(\hat{\mu}_i^2 + \hat{\mu}_j^2 + c_1) * (\hat{\sigma}_i^2 + \hat{\sigma}_j^2 + c_2)} \quad (6)$$

4.1 Experimental Observations on Quality Metrics

The section presents the outcomes achieved by applying the quality metrics on our proposed system. Table 1 depicts the outcomes obtained.

Table 1. Quality metrics for Ishihara filter enhanced images & histogram equalized images

Image No		1	2	3	4	5	6	7	8	9	10
Ishihara	PSNR	4.431	4.115	3.691	3.544	4.343	3.177	3.129	4.529	2.881	13.232
	MSE	0.360	0.384	0.427	0.442	0.368	0.047	0.481	0.486	0.352	0.515
	SSIM	0.315	0.239	0.202	0.116	0.323	0.763	0.165	0.152	0.239	0.124
Histogram	PSNR	0.581	5.470	4.497	6.390	6.908	5.133	4.122	4.097	6.023	3.308
	MSE	0.276	0.283	0.355	0.229	0.203	0.307	0.387	0.389	0.249	0.466
	SSIM	0.172	0.080	0.072	0.043	0.016	0.114	0.043	0.043	0.095	0.019

Analyzing the values depicted in Table 1, it is noticeable that histogram equalization yields a better result for a more significant number of images. However, for very few numbers of images Ishihara filter yields a better result.

4.2 Histogram Comparison

In this section, we have portrayed the comparative analysis of histogram plots for original image and histogram equalized image under different channels. Figure 6, 7, 8 depicts the same.

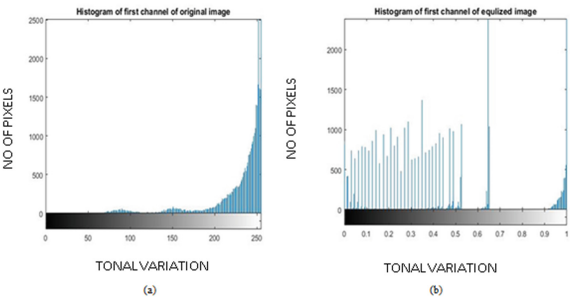


Fig. 6. Histogram comparison between (a) Original image & (b) Equalized image in the first channel

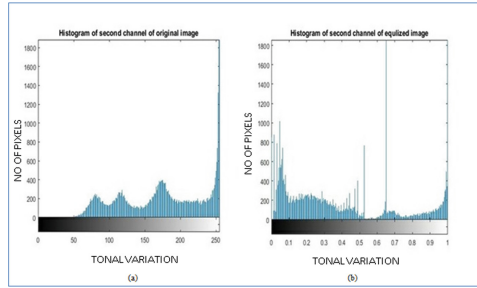


Fig. 7. Histogram comparison between (a) Original image & (b) Equalized image in second channel

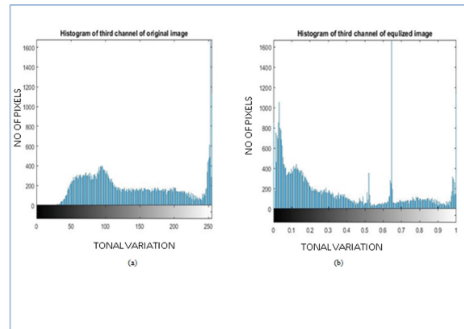


Fig. 8. Histogram comparison between (a) Original image & (b) Equalized image in the third channel

On analyzing the histogram plots of original image and enhanced image in different channels, it is evident that in every channel the tonal variation of histogram equalized image appears better than original image with uniform contrast.

4.3 Execution Time Analysis

Required execution time analysis is performed in this section to investigate the overhead involved in the Ishihara and Histogram equalization techniques. Required outcomes are depicted in Table 2.

Table 2. Execution time analysis

Image No	Execution time in seconds	
	Ishihara	Histogram equalization
1	5.6134	5.1291
2	5.6285	4.9457
3	5.7357	5.3145
4	5.7710	4.8507
5	7.2272	5.6992
6	6.0058	5.2433
7	5.9167	4.9630
8	5.9197	4.8883
9	5.7699	4.8840
10	6.102	5.723

Outcomes presented in Table 2 conforms that both the techniques consume almost same overhead. A narrow analysis proves that Histogram equalization is bit faster when compared to Ishihara.

4.4 Visual Comparison of Results Achieved

In this section, we have presented the visual analysis of the image generated from Ishihara filter and Histogram equalization in comparison with the original image. Figure 9 portrays gallery of images with the first image representing input, second image Ishihara generated and third image Histogram equalized.

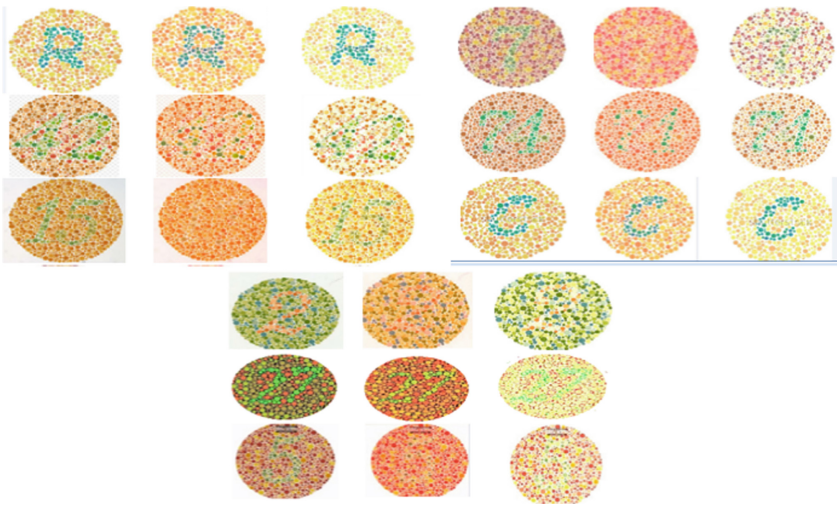


Fig. 9. Visual comparison of standard images

By analyzing Fig. 9, it evidently confirms that Histogram equalization works better for all images than Ishihara filter.

5 Conclusion and Future Work

A minimal amount of work was previously depicted for colour blind correction, which proposed implementing an Ishihara filter. But on the examination, it was found that the system failed to address the solution for some images. Hence our paper focused on improving the Ishihara filter through better enhancement and also proposed a histogram equalization as a parallel technique, providing better results for the images that were poorly addressed by the Ishihara filter. Outcomes proved that for most of the images, measured quality metrics for histogram equalization is better than Ishihara filter. Further investigation, even noted that the execution for the histogram equalization is bit faster than Ishihara filter. Hence this research paper concludes that the proposal has satisfied the requirements where a parallel technique for colour blind correction has yielded better results for the one that the Ishihara technique could not address. Therefore, rather than sticking to any of the specified techniques, as per the results produced by the paper, it is better to have the fusion of two different techniques to achieve a better result.

In future, research should focus on generalizing the outcomes that cover abundant studies and examine the execution of such processing for colour-blind people with a portable aid. Future work can also be concentrated on automatically automating the comparison to select an image with a better result.

References

1. Poret, S., Dony, R.D., Gregori, S.: Image processing for colour blindness correction. In: IEEE Toronto International Conference Science and Technology for Humanity (2009)
2. Martis, J.E., Sudeepa, K.B., Sannidhan, M.S., Bhandary, A.: A rapid automated process for organising bacterial cluster segments using deep neural networks. In: 2020 3rd International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 963–968. IEEE (2020)
3. Singh, D., et al.: Dknet: deep kuzushiji characters recognition network. IEEE Access **2022**, 1–12 (2022)
4. Venkataramu, A.C., Masahiko, A., Akshaya, A., Gurupura, A.P., Kumar, U., Rajashekar, R.R.: Offline signature recognition and verification using orb key point matching techniques. Adv. Sci. Technol. Eng. Syst. J. **5**(4), 01–07 (2020)
5. Sannidhan, M.S., Chaitra, K.M.: Assessment of image enhancement procedures for matching sketches to photos. In: 2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), pp. 1–5. IEEE (2019)
6. Alharbi, A.H., Aravinda, C.V., Shetty, J., Jabarulla, M.Y., Sudeepa, K.B., Singh, S.K.: Computational models-based detection of peripheral malarial parasites in blood smears. Contrast Media Mol. Imaging **2022**, 9171343 (2022)
7. Venkataramu, A.C., Devaraju, S., Nanjundaiah, P.H., Reddy, K.R.U.: Off-line signature confirmation based on clusterrepresentations of geometrical and statistical features through vector distance. Neural Netw. Support Vector (2022)
8. Nigam, P.K., Bhattacharya, M.: Colour vision deficiency correction in image processing. In: IEEE International Conference on Bioinformatics and Biomedicine (2013)

9. Lee, J., dos Santos, W.P.: Fuzzy based simulation of real colour blindness. In: Annual International Conference of the IEEE Engineering in Medicine and Biology (2013)
10. Sumathi, K., Anita, S., Himabindhu, C.: Image enhancement with histogram local minimal. In: IEEE International Conference on Power, Control, Signals, and Instrumentation Engineering (ICPCSI) (2017)
11. Goyal, P., Chaurasia, V., Meena, O.P.: Application of median filter in the removal of random valued impulse noise from the natural image. In: International Conference of Electronics, Communication, Aerospace technology (ICECA) (2017)
12. Mc Dowell, J.: Design of a colour sensing system to aid the colourblind. *IEEE Potentials* **27**(4), 34–39 (2008)
13. Utama, D.Q., et al.: Colour-blind test quantification using RGB primary colour cluster. In: International Conference on Information Technology Systems and Innovation (ICITSI) (2016)
14. Flatta, D.R., et al.: PRWEB: preservation of subjective responses to website colour schemas through automatic recolouring. In: CHI'13 Extended Abstracts on Human Factors in Computing Systems (2013)
15. Kulkarni, R.K., Lahoti, C.B., Mehar, S.: impulse denoising using improved progressive switching median filter. In: Proceedings of the International Conference and Workshop on Emerging Trends in Technology (2010)
16. Ananto, B.S., Sari, R.F., Harwahyu, R.: Colour transformation for colour-blind compensation on augmented reality system. In: International Conference on User Science and Engineering (i-USer) (2011)
17. Tu, L., Dong, C.: Histogram equalization and image feature matching. In: IEEE 6th International Conference on Image and Signal Processing (CISP), vol. 1, pp. 443–447 (2013)
18. Achar, S.B.: Comparative analysis of different image enhancement techniques for matching sketches to photos (2018)
19. Cheng, Y., Pedersen, M., Chen, G.: Evaluation of image quality metrics for sharpness enhancement. In: IEEE 10th International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 115–120 (2017)
20. Sannidhan, M.S., Prabhu, G.A., Robbins, D.E., Shasky, C.: Evaluating the performance of face sketch generation using generative adversarial networks. *Pattern Recogn. Lett.* **128**, 452–458 (2019)
21. Sannidhan, M.S., Ananth Prabhu, G., Chaitra, K.M., Mohanty, J.R.: Performance enhancement of generative adversarial network for photograph–sketch identification. *Soft Comput.* **27**, 1–18 (2021). <https://doi.org/10.1007/s00500-021-05700-w>
22. Fujikawa, Y., Li, H., Yue, X., et al.: Recognition of oracle bone inscriptions by using two deep learning models. *Int. J. Digital Humanit.* 1–15 (2022). <https://doi.org/10.1007/s42803-022-00044-9>
23. Yue, X., Li, H., Saho, K., Uemura, K., Aravinda, V., C, Meng, L.: Machine learning based apathy classification on doppler radar image for the elderly person. *Procedia Comput. Sci.* **187**, 146–151 (2021)



Color Image Cryptography Using Block and Pixel-Wise Permutations with 3D Chaotic Diffusion in Metaverse

Renjith V. Ravi¹ , Pushan Kumar Dutta² , and Sudipta Roy³

¹ Department of Electronics and Communication Engineering, M.E.A Engineering College, Malappuram, India

renjithravi@meaec.edu.in

² School of Engineering and Technology, Amity University, Kadampukur, Kolkata, India

pkdutta@kol.amity.edu

³ Artificial Intelligence and Data Science, Jio Institute, Navi 410206, Mumbai, India

sudipta1.roy@jioinstitute.edu.in

Abstract. All multimedia data types, including digital images, audio, text, and video, are now freely accessible online due to the phenomenal growth of internet technology solutions and wireless telecommunication systems, including internet and cellular networks. Cryptographic algorithms are needed to provide a certain level of security, dependability, and privacy and to keep sensitive information from being accessed. At the same time, data is being collected and sent. This paper presents a novel and effective system of color image cryptographic algorithms based on image pixel shuffling, which involves repeatedly splitting the image into a predetermined number of blocks and sub-blocks and spinning each clockwise by 90°. By obscuring the statistical makeup of the pixels, two dimensions of Arnold cat mapping were then used to distort further the connection between neighboring pixels in the plaintext image. The image pixels are returned to their original places during the decryption process. The Henon mapping algorithm is employed for image encryption to decrease the correlation between the ciphertext and the plaintext image. The encryption above algorithm was tried and examined on a few color images, and the results demonstrated a sufficient level of security to withstand differential and statistical attacks.

Keywords: Image encryption · Image cryptography · 3D Henon Map · Arnold shuffling · Metaverse

1 Introduction

The confidentiality of information during communication has become a crucial and pressing problem as a result of the widespread transmission of multimedia data through wireless networks and the internet, particularly for digital images owing to their advantageous characteristics, such as their high data capacity, high levels of redundancy, and high correlation between individual image data [1]. Multimedia encryption is a necessary approach for secure image and video data transmission. A primary objective of

an image cryptographic algorithm is to transform readable image data into unreadable data. This safeguards the original image. Digital image encryption may be found in a wide variety of applications, including the military, archaeology applications, medical images, video surveillance, secret transfer, and everyday activities such as financial data. Researchers were given a number of digital algorithms for image encryption among this prominent multimedia data, all of which function to prevent unauthorized access to the image's information on the network. Many solutions, including chaos-based solutions, data processing-based techniques, key generation-based cryptosystems, and so on, have been presented in recent years for the real-time security of image transmission through modern communication networks. One such system is chaotic and relies on diffusion and confusion processes. It performs better in cryptographic systems due to the properties such as pseudo-random behavior, sensitivity to the starting state, a non-linear dynamic nature, and unexpected ways.

This study suggests a block permutation, diffusion, and pixel position permutation-based image encryption technique. The remaining portion of this study is organized as follows: Works pertaining to image encryption are presented in Sect. 2. In Sect. 3, an image encryption technique based on the 2D Arnold cat map (ACM) and 3D Henon chaotic map (HCM) is suggested and analysed. We also talk about the key feature of the Arnold cat map and an image-based scrambling technique. Section 4 discusses the experiment findings, and the proposed method's security is evaluated using formal verification. The conclusion of the article was finally covered in Sect. 5.

2 Literature Survey

Some studies [2] focus specifically on the applicability of chaos theory in image encryption. However, a few of the studies examined above, such as the one described in [3], are too old. Furthermore, although some provide a roadmap for the future, none succeeds in creating an infrastructure for chaotic cryptography. A deliberately altered medical image, according to Deepa and Sivamangai [4], makes it more challenging to identify a real condition. The secrecy of clinical images is now more important than ever. On the other side, the encryption procedure might significantly burden the systems used for medical processing applications. They asserted that DNA cryptography and chaos cryptography are the best methods for resolving this trade-off. They provided quantitative and qualitative data taken from previously published, relevant research papers to demonstrate how the technologies above handle the tradeoff.

Additionally, they developed a few criteria for future study in this field. Yadav and Chaware [5] contend that there are flaws in the current encryption and information-hiding systems that make it possible for data to be stolen and copyrights to be violated. They started by reviewing current image encryption techniques, focusing particularly on cooperative encoding in cryptography. Then, with the assistance of the AES and substitution boxes, they suggested a revolutionary technique utilizing LDPC code and chaotic S-boxes. A comparative perspective is used in certain existing assessments.

For instance, [6] examined the benefits and drawbacks of the various chaotic image encryption techniques. In [7], the authors analyzed and contrasted various unidimensional chaotic systems with some maps with hyper-dimension with regard to their functionalities in cryptography. This is another pertinent survey that was presented. Another

example is chaotic encryption, which the authors of [8] cited as a potential method for encrypting images and movies in which the correlation between neighboring pixels is strong. In order to choose the best chaotic map, they gave an analysis of the current chaotic algorithms for image encryption. After evaluating logistic maps, tent maps, sine maps, and other chaotic maps, ACM was identified as the most appealing map for this application. Additionally, in [9], the authors compared and contrasted the effectiveness of image encryption techniques based on the five conventional techniques AES, DES, RSA, and Blowfish with certain chaos-based techniques.

3 Materials and Methods

3.1 3D Henon Map

The 3D Henon map [10, 11] is an example of a straightforward quadric autonomous system that spans three dimensions and will be characterized by the equation

$$\begin{aligned}x_{i+1} &= a - y_i^2 - bz_i \\ y_{i+1} &= x_i \\ z_{i+1} &= y_i\end{aligned}\tag{1}$$

The process is discrete across time, and (x_i, y_i, z_i) , where i stand in for the iteration number, depicts its state. Any subsequent iteration $i + 1$ computes the state $(x_{i+1}, y_{i+1}, z_{i+1})$ using (x_i, y_i, z_i) . a and b are both 1.76 and 0.1 in this note. The starting value of the chaotic map would be a collection of constant values (x_i, y_i, z_i) chosen from the interval $[-1.18, 1.85]$ for each output stage $(x_{i+1}, y_{i+1}, z_{i+1})$.

The 3D Henon map has a basic geometric understanding of $i + 1$ and is reversible. Additionally, because its coefficient is an integer, it will reduce calculating errors.

3.2 Pixel Permutation Using Arnold Cat Map

The Arnold's Cat Map (ACM) transformation is used to scramble the pixels of a color image and achieve increased cypher system security. The 2D ACM does not modify the intensity value of each pixel; instead, it shuffles the image's data.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & p \\ q & pq + 1 \end{bmatrix}^{-1} \times \begin{bmatrix} x' \\ y' \end{bmatrix}\tag{2}$$

After multiple repetitions of applying the 2D Arnold cat transformation, the connection between neighboring pixels is completely destroyed, and the plaintext image seems deformed and unintelligible [10, 11]. If you iterate anything enough times, it will eventually revert to its initial appearance. This indicates that the ACMS is a periodic transformation. Image shuffling ensures that the statistical characteristics of the encrypted image and the plaintext image are the same, strengthening the cryptosystem's integrity. An example of a simple image and a scrambled image with 6 iterations are shown, and their respective grayscale histograms are shown in Fig. 1.

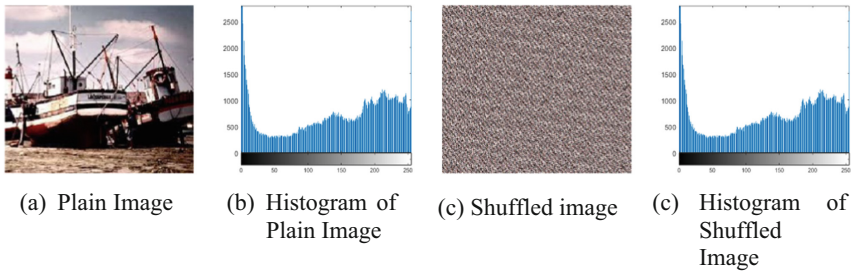


Fig. 1. Example of shuffled image and its histogram

3.3 Block Operation

When commencing the encryption and decryption operations, the image’s histogram remains steady because RGB image scrambling makes it possible to disrupt the correlation between adjacent pixels by altering their location rather than their value. Three processes are involved in moving the image pixels.

Start by dividing the image into four components, known as quadrants and rotating each one 90° in the direction of the clock. Then, create 16 image blocks by splitting each quadrant into four smaller quadrants and rotating them 90°. The next step is to split each block into four pieces and rotate each piece 90° to create 64 image blocks [1] (Fig. 2).

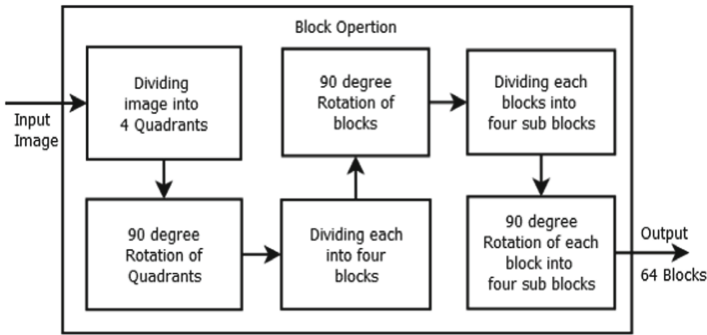


Fig. 2. Block operation

3.4 Proposed Algorithm

Image cryptography’s main objective is to ensure that all authorized users can use a simple, inexpensive way to encrypt and decrypt data. The primary algorithm used in this study is shown in Fig. 1. It comprises encryption and decryption procedures, utilizing a three-dimensional pixel matrix RGB image with dimensions of $256 \times 256 \times 3$.

An innovative method for quick and safe digital image ciphering is presented in this study. The first step of the suggested system is to jumble the location of the pixels

by repeatedly breaking the two-dimensional image data into blocks and spinning each block 90° in the clockwise direction for decorrelating the relationships between pixels. Second, Arnold cat mapping is a good fit for this item to improve security and provide more pixel shuffle in the recommended encrypted approach. The image is then encrypted using Henson's chaotic technique. The chaotic mapping approach is intended to blur the relationship between the actual plaintext and the encrypted ciphertext images. The main goal of the suggested techniques is to avoid losing image data during the decryption, making it considerably simpler to reconstruct the plain image from the cypher image (Fig. 3).

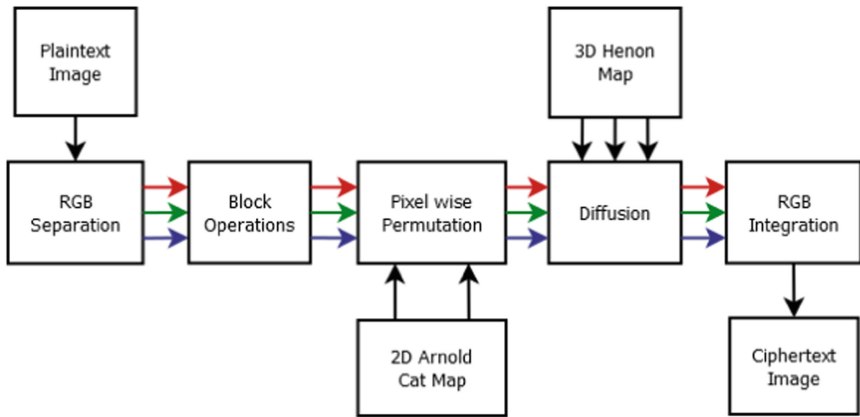


Fig. 3. Proposed encryption approach

3.5 Key Generator Using 3D Henon Map

We use a matrix of dimensions $(256 \times 256 \times 3)$ to represent the RGB colour image, where 256 denotes the colour level for both the rows and columns. To encrypt a colour image, we must divide the RGB matrix into its individual parts and combine each R, G, and B matrix image into a new array comprising (1×65536) components. In order to create the key sequence for the encryption system, we first create a matrix with a dimension of (1×193668) components using the Henon map of Eq. 1. The two secret values, a and b , are expressed as a secret key to create the following keys, with starting values of $x_0 = 0.01$ and $y_0 = 0.02$. The $[-11]$ interval contains the sequence's critical values. We must convert [1] them into numbers in the range $[1256]$ using Eq. 3.

$$k_{new} = (k_{old} \times 10^3) \bmod 256 \quad (3)$$

Image encryption is accomplished by the process of diffusion, which replaces the jumbled image blocks with new ones. This is done by altering the value of each component of the R, G, and B pixels by performing a bitwise XOR operation with the key attributes that are designated for each component.

3.6 The Decryption of Ciphertext Image

We employ the inverse process of the encryption stages to reconstruct the plaintext image. Due to the unpredictable nature of the chaotic system attribute, the original image is constructed using the same key value sequences as the scrambled image. This is done by using an exclusive function that combines the cypher image with the original image. Then, the original image is made by rearranging the blocks of the scrambled image by turning them 90° counterclockwise. The inversion of Arnold mapping may be used to create the spinning image blocks.

4 Experiments and Results

In order to determine how different, the cypher images are from the original images, we performed a number of statistical and security analyses, including an analysis of histogram for both plaintext and ciphertext images, a number of pixels change rate (NPCR) analysis, a unified average changing intensity (UACI) analysis, a key estimation, and a correlation coefficients assessment (Fig. 4).

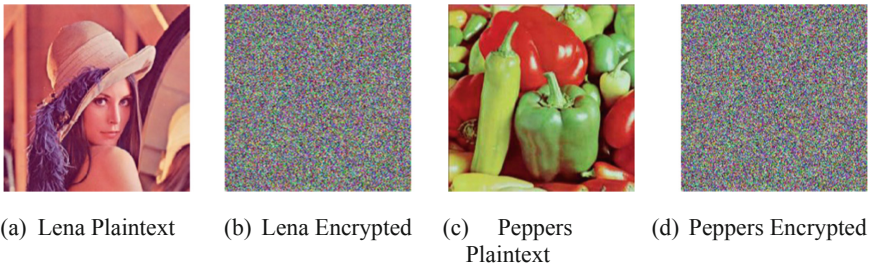


Fig. 4. Results of encryption and decryption

4.1 Statistical Attack Analysis

The entire R, G, and B colour components of the plaintext image are subjected to a histogram in order to evaluate the efficacy of the suggested encryption approach and its stability against statistical assaults. Figure 5 displays the histograms of test images using the suggested approach, both plaintext and ciphertext. From all the images, it is clear that there is a perceptible difference between the plaintext image and the encrypted image pixels' frequency values and the visual depiction of all the color channels' histograms. In light of the fact that histogram criteria cannot provide any information to statistical cryptanalysis for deciphering the encryption algorithm, so the proposed method is an excellent technique for masking any traces of the original image.

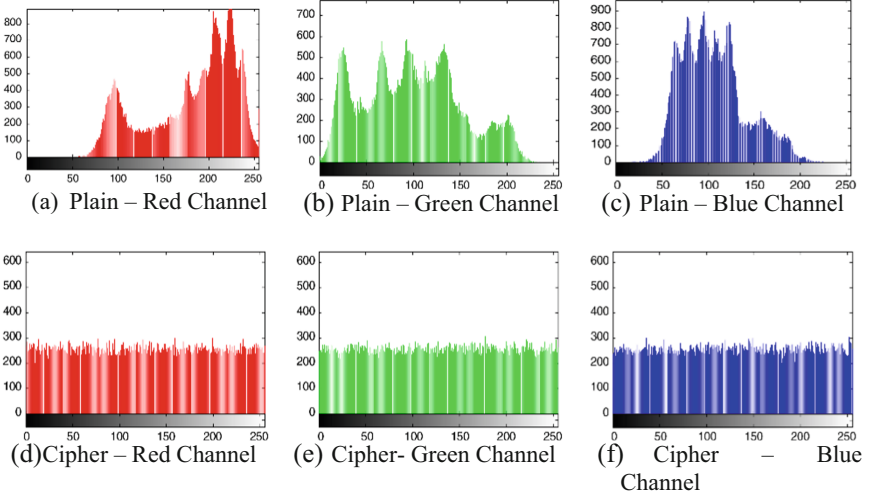


Fig. 5. Histograms of Red, Green and Blue components of plaintext and ciphertext of test image *Lena*

4.2 Correlation Coefficient Analysis

A plaintext and ciphered image pixel's behavior, which is completely random and extremely uncorrelated in its horizontal, major diagonal, vertical, and horizontal orientations, is completely concealed by a good encryption technique, according to any image encryption scheme. As is well known, the correlation coefficient (CC) between neighboring pixels in a ciphertext image is a useful statistic for assessing the encryption strength of any cryptosystem. Three variables are needed to determine the correlation coefficient, and they are as in Eqs. 4, 5, 6, and 7:

$$\text{cov}(u, v) = \frac{1}{N} \sum_{i=1}^N (x_i - E(u))(y_i - E(v)) \quad (4)$$

$$D(u) = \frac{1}{N} \sum_{i=1}^N (u_i - E(u))^2 \quad (5)$$

$$E(u) = \frac{1}{N} \sum_{i=1}^N u_i \quad (6)$$

$$r_{xy} = \frac{\text{cov}(u, v)}{\sqrt{D(u)}\sqrt{D(v)}} \quad (7)$$

Here, u and v stand for the colour saturation of two adjacent pixels of the original or cypher image. The autocorrelation for both plaintext and cipher images is shown in Table 1.

Table 1. Correlation coefficients

Image	Color plane	Direction		
		Horizontal	Vertical	Diagonal
Lena Plaintext	Red	0.9456328	0.9512843	0.8935267
	Green	0.9748237	0.9744362	0.9483654
	Blue	0.9384260	0.9268356	0.8731619
Lena Ciphertext	Red	− 0.0003568	− 0.0007836	0.0003247
	Green	− 0.0007184	− 0.0007956	− 0.0001356
	Blue	− 0.0007684	− 0.0001579	− 0.0001762

The value of CC will be close to one of the corrected pixels that are highly linked. If the coefficient, on the other hand, is near 0, the pixels are not connected. The auto-correlation for the cipher image is clearly shown to be extremely low and close to zero. This proves that the proposed cryptographic algorithm produces better and more secure encryption.

4.3 Analysis of Differential Attack

The NPCR and the UACI are the measurements used to assess the distinctions between the encrypted and decrypted images. The mathematical expression for the NPCR measure is provided in Eq. (8) and Eq. (9).

$$\text{NPCR} = \frac{1}{MN} \sum_i^M \sum_j^N D(i, j) \times 100\% \quad (8)$$

$$D(i, j) = \begin{cases} 1, I_1(i, j) \neq I_2(i, j) \\ 0, I_1(i, j) = I_2(i, j) \end{cases} \quad (9)$$

where $I_1(i, j)$ denotes the encrypted image and $I_2(i, j)$ denotes the original image. When just one pixel from the input images is modified, the NPCR value represents the changed average number of pixels in the ciphertext image.

The intensity rate of variations between the input images and the encrypted image is measured by UACI. Equation (10) displays the UACI measure's mathematical formulation (Table 2).

$$\text{UACI} = \frac{1}{MN} \sum_i^M \sum_j^N \frac{|I_1(i, j) - I_2(i, j)|}{255} \times 100\% \quad (10)$$

Table 2. Values obtained for NPCRF and UACI

Image	NPCR			UACI		
Lena	99.9146	99.6833	99.8223	33.8337	33.4483	33.6023
Peppers	99.8283	99.7897	99.8114	33.4613	33.4835	33.7224

5 Conclusion

This article suggests a color image ciphering method based on a chaotic 2D Henon-like system. The suggested technique included two steps of pixel location and pixel jumbling using confusion techniques. First, 64 blocks of the original image were divided, and each block was rotated 90° clockwise. Second, add extra pixel scrambling to blocked images using 2D Arnold Cat mapping. In order to encrypt the scrambled picture, diffusion and shuffling methods are used to alter the grey levels of the pixel intensities. The chosen channel passes through a 3D Henon map to generate a series of random values that combine with the pixel value of the preferred channel that has been shuffled to create the encrypted image and increases its resistance to all attacks, including statistical and differential attacks. Any RGB component may be used with the proposed cryptosystem technique. The results of the experiments and the numerical analysis demonstrate that the proposed cryptosystem is suitable for efficient image encryption and attains more adaptability, dependability, and encryption quality.

References

1. Alzubaidi, A.M.N.: Color image encryption and decryption using pixel shuffling with henon chaotic system. *Int. J. Eng. Res.* **3** (2014)
2. Zolfaghari, B., Koshiba, T.: Chaotic image encryption: state-of-the-art, ecosystem, and future roadmap. *Appl. Syst. Innov.* **5**, 57 (2022)
3. Sankpal, P.R., Vijaya, P.A.: Image encryption using chaotic maps: a survey. In: 2014 5th International Conference on Signal and Image Processing (2014)
4. Deepa, N.R., Sivamangai, N.M.: A state-of-art model of encrypting medical image using DNA cryptography and hybrid chaos map-2d zaslavski map. In: 2022 6th International Conference on Devices, Circuits and Systems (ICDCS) (2022)
5. Yadav, K., Chaware, T.: Review of joint encoding and encryption for image transmission using chaotic map, LDPC and AES encryption. In: 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC) (2021)
6. Suneja, K., Dua, S., Dua, M.: A review of chaos based image encryption. In: 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) (2019)
7. Bu, Y.: Overview of image encryption based on chaotic system. In: 2021 2nd International Conference on Computing and Data Science (CDS) (2021)
8. Ayad, J., et al.: Image encryption using chaotic techniques: a survey study. In: 2021 International Conference in Advances in Power, Signal, and Information Technology (APSIT) (2021)
9. Thein, N., Nugroho, H.A., Adji, T.B., Mustika, I.W.: Comparative performance study on ordinary and chaos image encryption schemes. In: 2017 International Conference on Advanced Computing and Applications (ACOMP) (2017)
10. Sava, D., Vlad, A., Tataru, R.: A new type of keystream generator based on chaotic maps: illustration on a Hénon generalized map. In: 2014 10th International Conference on Communications (COMM) (2014)
11. Albhrany, D.E.A., Alshekly, T.: A new key stream generator based on 3d Henon map and 3d cat map. *Int. J. Sci. Eng. Res.* **8** (2017)

Sentiment Analysis and Recommendation Systems



Review on Recent Trends in Recommender Systems for Smart Cities

Sana Abakarim^(✉) , Sara Qassimi , and Said Rakrak 

L2IS Laboratory, Faculty of Science and Techniques, Cadi Ayyad University,
Marrakesh, Morocco

sana.abakarim@ced.uca.ma, sara.qassimi@uca.ac.ma, s.rakrak@uca.ma

Abstract. Smart cities are defined as municipalities that use ICT to consider the requirements of its urban community. Indeed, the emergent technologies enable to improve the quality of life of citizens, the quality of public services, and to increase information-sharing with the public. Within ICT, recommender systems are among the most popular and powerful tools for filtering information, improving relationships between government and civil society stakeholders, and supporting decision-making through technology platforms. In this paper, we will highlight the presence and importance of using recommender systems in smart cities. Our survey highlights current trends and challenges to enhance decision-making for customers, businesses, and government agencies by using personalized recommendations.

Keywords: Recommender systems · Smart cities · Interconnected city · Intelligent city · Sensors · Internet of Things · Open data

1 Introduction

Smart Cities term was first introduced in the 1990s, due to the use of Information and Communications Technologies (ICT) for a city's modern infrastructure. Afterwards, the idea of the smart city has continued to evolve, and today is a vague concept and is not always used consistently [1]. Some authors consider that a smart city is a city where ICT is highly present and applied to critical infrastructure components and services [2]. The term "smart city" refers to an "instrumented, interconnected and intelligent city" [3], the term 'instrumented' refers to the ability to capture and integrate real-world live data by the use of sensors, software applications, cameras, personal devices and other similar sensors. In this context, sustainability includes reduced use of non-renewable resources, environmental conservation, diverse and robust economy, independent communities, economic vitality and diversity, autonomy in community, citizen well-being, and satisfaction of basic human needs [4].

Critics have argued that the design of online public services was too often focused on technological possibilities rather than user needs, and thus called for

more user-centered services [5]. Therefore, to address this issue, progress must be made in developing personalized approaches that not only improve service effectiveness and efficiency, but also increase user satisfaction [6]. In addition, the huge amount of data and services also requires the implementation of information filtering approaches that aim to solve existing problems in a city, targeting its various stakeholders. It is in these contexts where recommender systems (RS) have challenging opportunities [7]. In particular, in smart cities, recommender systems can be used to improve relationships between stakeholders, e.g., governments and civil society, and assist in decision-making in the city through technology platforms.

There are many research studies that cover the application of recommender systems in smart cities. However, to the best of our knowledge, there is no comprehensive review of recent trends in recommender systems for smart cities. Motivated by this fact, in this paper we present a review of the state of the art on recent trends of RS for smart cities, as well as providing an appropriate guiding framework for future research.

The paper is organized as follows. Section 2 presents recommender systems approaches and applications. Section 3 explains the dimensions in which smart cities are commonly classified. Section 4 presents recommender systems and its challenges. Section 5 delivers the conclusion.

2 Recommender Systems

Recommender systems are defined as an information filtering system designed to aid decision making as an intuitive line of defense against consumer over-choice [8]. In order to provide personalized suggestions, recommender systems make use of past choices and preferences to reflect the users' tastes and interests (information can be either explicitly provided or implicitly inferred), other features that can be used to model users such as demographic data, personality traits, emotional states or trust relationships [9].

There are many ways to build RS, such as using modeling-centric approaches like collaborative filtering, content based, or hybrid systems [10]. Content based Recommender system is a method of recommending similar items based on information about items selected by the user in the past [11]. Collaborative filtering is a model that builds a database of user preferences using user ratings data to predict items that match the user's tastes, and then uses it for recommendation [13]. A hybrid recommender systems use a combination of two or more recommender approaches in different ways to achieve their complementary benefits [26].

RS have undergone a number of recent evolutions, and as a result, the number of studies is increasing every year. The year distribution of the studies is presented in Fig. 1. It shows that articles on RS have been increasing rapidly in recent years.

In recent years, Rs have been widely used in many fields, such as streaming services, education services, e-commerce services, tourism services, social networking services, healthcare services and academic services [12]. The distribution of studies by application services is presented in Fig. 3. [20] propose a

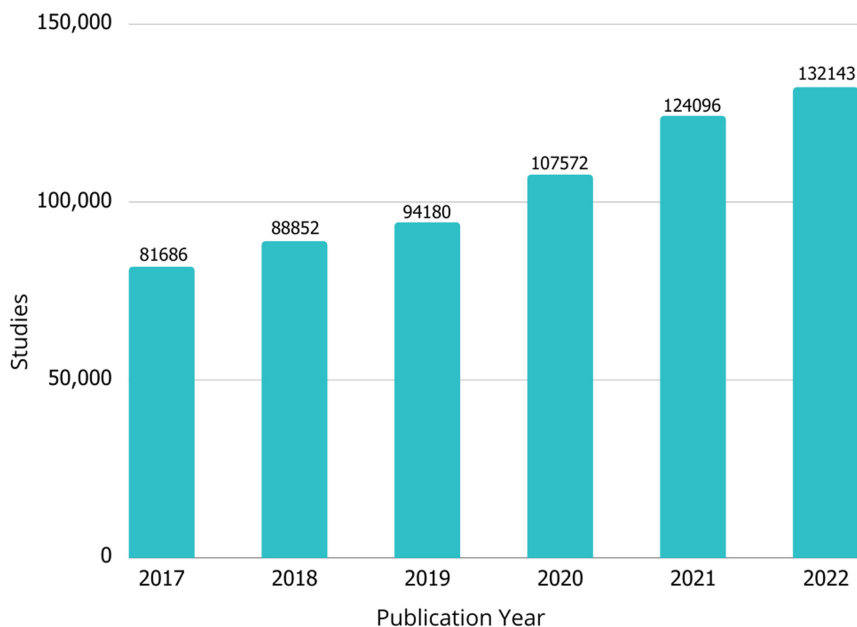


Fig. 1. Distribution of studies by publication year.

recommendation system called MOREc, which produces Top-N music recommendations by priming the system with previously collected data. For [15], they present a solution named AyudasCBI, which is a system that uses RS to propose personalized activities to students to enhance their individualized education. In the e-commerce fields, the authors of [18], suggest a hybrid model that manages both user and article metadata in order to solve the sparsity problem. [16] offer a personalized list of tourist attractions for each tourist based on the similarity of their interests.

3 Recommender System and Smart Cities

There are various definitions of smart city. The lack of unified definitions of smart city is due to the different entities involved and the functionality of the smart city. The synthetic definition of a smart city, according to [14], is that to become a smart city, it is necessary to connect multiple data sources from a range of urban activities and domains to reveal opportunities to deliver innovations to the connected citizens of today. There is a number of relevant dimensions where smart city initiatives can be classified. One of many definitions, [17] has proposed six dimensions: smart economy, smart environment, smart mobility, smart governance, smart living, and smart people (See Fig. 2).

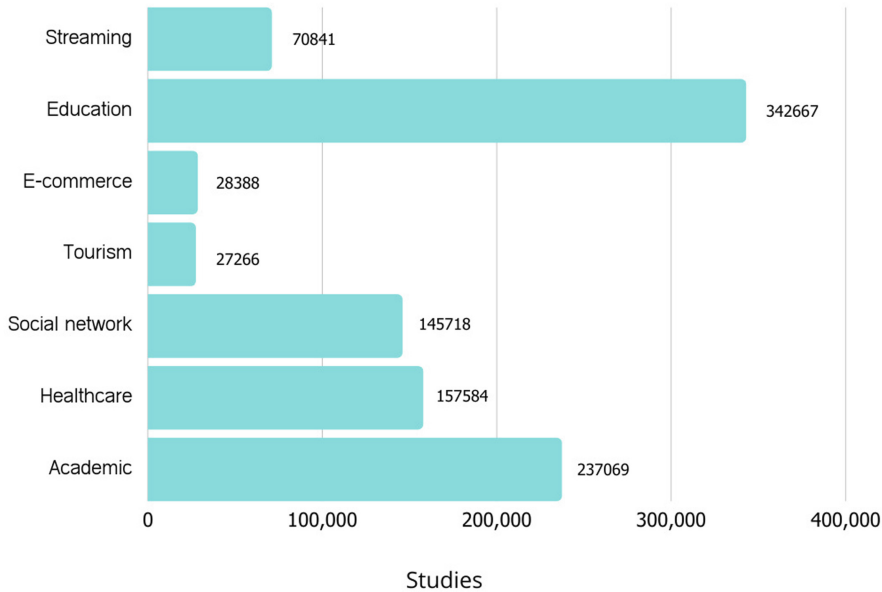


Fig. 2. Distribution of studies by application services.

Detractors have claimed that the creation of online public services is too often focused on technological possibilities instead of user needs and have therefore called for more user-centered services [5]. In response, the development of more user-centric online services is more important. Therefore, progress must be made in developing personalized approaches that not only make online services more effective but also increase user satisfaction [6]. In this context, recommender systems have proposed to provide new stimulating opportunities.

Several studies have been done using recommender systems in the different dimensions mentioned (smart economy, smart environment, smart mobility, smart governance, smart living, and smart people). These studies are presented in the following subsections.

3.1 Recommender Systems for Smart Economy

In this dimension, smart solutions aim to encourage economic growth through various actions that aim to promote labor market flexibility, productivity and transformation capacity [17]. The smart economy dimension has three main objectives: to promote local businesses, to improve economic productivity, and to support business and trade networks in a city.

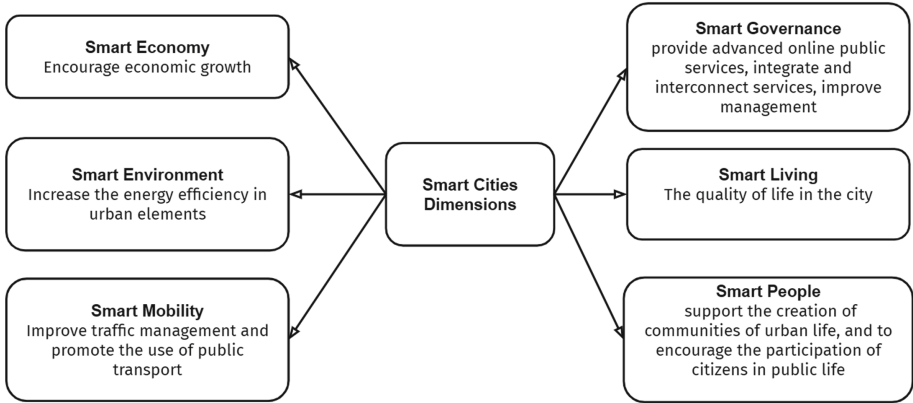


Fig. 3. Smart cities dimensions.

Promoting Local Businesses: Concerning the economic development of a city, [19] proposed a system that coordinates stakeholders, communicates and produces automatic stock updates, provides healthy eating recommendations to citizens, and suggests healthy menus to restaurants based on local farmers' stock.

Improving Economic Productivity: In order to speed up delivery and optimizing carrier time and reducing travel costs, [21] proposed a collaborative logistics platform to standardize delivery requests and reduce the number of trucks needed. On the other hand [22] presented TaxiRec, a system that evaluates and recommends groups of routes with a large number of passengers.

Supporting Business and Commerce Networks: [23] proposed a recommender system application called Smart city social Agent Network, which allows for the presence of groups of social agents, formed on the basis of trust measures, and the global topology derives from the integration of these groups, which are dynamically adapted over time to the changing desires and needs of users (Table 1).

3.2 Recommender Systems for Smart Environment

The "smart environment" dimension aims to increase the energy efficiency in urban elements, such as homes and buildings, and to manage a city's environmental resources, such as water, air quality, waste and food. To boost environmental and economic durability, [24] presented a smart closet prototype that recommends what to donate or recycle.

3.3 Recommender Systems for Smart Mobility

One of the main goals in the context of mobility in a smart city, is to improve traffic management and promote the use of public transport, and non-motorized

Table 1. Recommendation system features of the studied articles in different services.

Service	Objective	Data source	Literature sources	Recommended items
Smart economy	Promote local businesses	IOT	[19]	Food to offer or to eat
	Improve economic productivity	Databases	[21]	Locations to deliver packages
		GPS	[22]	Road clustering
	Support business and trade networks	–	[23]	IoT services
Smart environment	Increasing recycling	IOT, Databases	[24]	Clothes for recycling
Smart mobility	Improve traffic management	Sensors	[25]	Driving roads
	Promote the use of public transport	IOT, Sensors	[27]	Best road less traveled
	Promote non-motorized and clean mobility options	Sensors	[28]	Whom to travel with
Smart governance	Provide advanced online government services	web services	[29]	Government services
		Questionnaires	[29]	Politic candidates
Smart living	Provide accessibility facilities	Databases, Expert knowledge	[32]	Software
	Support health measures	Sensors	[33]	Media Use Guidelines
Smart people	Encourage citizen participation in public life	Web platforms	[34]	Interesting projects
		Social media	[35]	Local Events & News

and clean mobility options, to encourage citizens to use more sustainable means of transportation.

Enhancing the Traffic Management: In order to reduce traffic congestion, [25] provides a solution that recommends whether a given vehicle should change its route or not. While to address parking management problems, [25] proposed parking spaces recommendations.

Promoting the use of Public Transport: Addressing this issue [27] encourages the use of public transport by improving the user experience.

Promoting Non-motorized and Clean Mobility Options: In this context, [28] suggested a vehicle sharing recommendation system that recommends users with whom to travel.

3.4 Recommender Systems for Smart Governance

This dimension aims to provide advanced online public services, integrate and interconnect services, improve management, regulation and delivery of services, and enable citizens to collaborate in creating and improving services. In this context, several applications are proposed, to provide citizens with personalized government e-notifications and e-services [29], to keep the government informed of citizens' issues, concerns and opinions expressed on e-consultation and e-participation platforms [30], to help find business partners in government e-services [31], to provide government employees with personalized business event recommendations [7].

3.5 Recommender Systems for Smart Living

The “smart life” dimension includes aspects related to the quality of life in the city, recommender system on this dimension addresses the following objectives: provide accessibility facilities, Support health measures.

Provide Accessibility Facilities: In this domain [32] proposed a system that helps people with disabilities install assistive software (AS) in devices/controllers.

Support Health Measures: To promote healthy lifestyle and well-being, [33] provided a system that offers personalized media recommendations, and various alerts for elderly and dependent individuals.

3.6 Recommender Systems for Smart People

The main objectives of the “smart people” actions are to support the creation of communities and the management of urban life, and to encourage the participation of citizens in public life by using ICT. Recommender systems in the “smart people” dimension aim to support community building and urban management, with the objective of enabling urban residents to better enjoy community life. For example, [34] developed a system for sharing information about community projects with citizens. The recommendation system also seeks to promote a participatory society. For instance, the interaction between citizens is enabled by recommending information from social media and information providers [35].

4 Challenges and Open Issues

Recommender systems and smart cities in general face several challenges, these challenges are primarily in data processing and storage, where the rapid increase

in the number of sensors and devices deployed is rapidly generating huge amounts of data of different types (structured and unstructured), this challenge, considered as a Big Data challenge, is addressed on the one hand by collecting, organizing and processing the data, and on the other hand by verifying the quality and reliability of the data [26]. Privacy is also considered one of the main challenges, the numerous uses of sensors and cameras installed in various places (streets, restaurants, buses, parking lots, etc.), give people the impression that they are being spied on by the government, and that their private and personal life is in danger. To solve the trust problem, governments and legislators must adopt privacy laws and policies that regulate the use of individuals' personal information and increase transparency [36]. Scalability is another challenge of recommender systems and smart cities, due to the rapid growth of the Internet of Things, scalability is a critical issue that must be addressed at several levels, such as security level, routing and network level, and resource constraints. Using a cloud platform as a centralized compute and storage resource offers many benefits and solves the resource limitation problem [36].

5 Conclusion

In this paper, we highlight the presence and importance of the recommender system in smart cities, presenting different application areas and challenges. This study revealed that the main smart city actions for which recommender systems have been proposed are energy efficiency, notification of government e-services, promotion of healthy living, optimization of parking space utilization, and reduction of traffic congestion. The recommendation system in smart cities still presents many challenges, including privacy, scalability, and data processing and storage.

References

1. Albino, V., Berardi, U., Dangelico, R.M.: Smart cities: definitions, dimensions, performance, and initiatives. *J. Urban Technol.* **22**(1), 3–21 (2015). <https://doi.org/10.1080/10630732.2014.942092>
2. Washburn, D., Sindhu, U., Balaouras, S., Dines, R.A., Hayes, N.M., Nelson, L.E.: Helping CIOs Understand “Smart City” Initiatives: Defining the Smart City, Its Drivers, and the Role of the CIO. Forrester Research Inc., Cambridge (2010)
3. Harrison, C., et al.: Foundations for smarter cities. *IBM J. Res. Develop.* **54**(4), 1–16 (2010). <https://doi.org/10.1147/JRD.2010.2048257>
4. Bouzguenda, I., Alalouch, C., Fava, N.: Towards smart sustainable cities: a review of the role digital citizen participation could play in advancing social sustainability. *Sustain. Cities Soc.* **50**, 101627 (2019). <https://doi.org/10.1016/j.scs.2019>
5. Verdegem, P., Verleye, G.: User-centered E-government in practice: a comprehensive model for measuring user satisfaction. *Gov. Inf. Q.* **26**, 487–497 (2009). <https://doi.org/10.1016/j.giq.2009.03.005>
6. Dawes, S.S.: The evolution and continuing challenges of e-governance. *Adm. Rev.* **68**, S86–S102 (2008). <https://doi.org/10.1111/j.1540-6210.2008.00981.x>

7. Cortés-Cediel, M., Cantador, I., Gil, O.: Recommender systems for e-governance in smart cities: state of the art and research opportunities. In: Proceedings of the 1st International Workshop on Recommender Systems for Citizens. ACM, vol. 7 (2017). <https://doi.org/10.1145/3127325.3128331>
8. Zhang, S., Yao, L., Sun, A., Tay, Y.: Deep learning based recommender system: a survey and new perspectives. *ACM Comput. Surv.* **52**(1), 38 (2019). <https://doi.org/10.1145/3285029>
9. Quijano-Sanchez, L., Cantador, I., Cortés-Cediel, M., Gil, O.: Recommender systems for smart cities. *Inf. Syst.* **92**, 101545 (2020). <https://doi.org/10.1016/j.is.2020.101545>
10. Najmani, K., El habib, B., Sael, N., Zellou, A.: A comparative study on recommender systems approaches. In: Proceedings of the 4th International Conference on Big Data and Internet of Things (BDIoT'19). Association for Computing Machinery, New York, NY, USA, Article 3, pp. 1–5 (2020). <https://doi.org/10.1145/3372938.3372941>
11. Javed, U., Shaukat, K., Hameed, I.A., Iqbal, F., Alam, T.M., Luo, S.: A review of content-based and context-based recommendation systems. *Int. J. Emerg. Technol. Learn. (iJET)*, **16**(3), 274–306 (2021). Kassel, Germany: International Journal of Emerging Technology in Learning. Accessed 23 Nov 2022. <https://www.learntechlib.org/p/219036/>
12. Ko, H., Lee, S., Park, Y., Choi, A.: A survey of recommendation systems: recommendation models, techniques, and application fields. *Electronics* **11**, 141 (2022). <https://doi.org/10.3390/electronics11010141>
13. Im, I., Hars, A.: Does a one-size recommendation system fit all? the effectiveness of collaborative filtering-based recommendation systems across different domains and search modes. *ACM Trans. Inf. Syst.* **26**(1), 4 (2007). <https://doi.org/10.1145/1292591.1292595>
14. Lai, C.S., et al.: A review of technical standards for smart cities. *Clean Technol.* **2**(3), 290–310 (2020). <https://doi.org/10.3390/cleantechnol2030019>
15. Tejeda-Lorente, Á., Bernabé-Moreno, J., Porcel, C., Galindo-Moreno, P., Herrera-Viedma, E.: A dynamic recommender system as reinforcement for personalized education by a fuzzily linguistic web system. *Procedia Comput. Sci.* **55**, 1143–1150, (2015). <https://doi.org/10.1016/j.procs.2015.07.084>. ISSN: 1877-0509
16. Esmaeili, L., Mardani, S., Golpayegani, S.A.H., Madar, Z.Z.: A novel tourism recommender system in the context of social commerce. *Expert Syst. Appl.* **149**, 113301 (2020). <https://doi.org/10.1016/j.eswa.2020.113301>. ISSN: 0957-4174
17. Giffinger, R., Fertner, C., Kramar, H., Meijers, E.: City-ranking of European medium-sized cities. *Centre Reg. Sci. Vienna Univ. Technol.* **2007**, 12 (2007)
18. Khan, Z., Hussain, M.I., Iltaf, N., Kim, J., Jeon, M.: Contextual recommender system for E-commerce applications. *Appl. Soft Comp.* **109**, 107552 (2021). <https://doi.org/10.1016/j.asoc.2021.107552>. ISSN: 1568-4946
19. Dennis, A.L.R., Ahrary, A., Horibe, N., Yang, W.S.: IoT-security approach analysis for the novel nutrition-based vegetable production and distribution system. In: 3rd IIAI International Conference on Advanced Applied Informatics, IIAI-AAI 2014, Kitakyushu, Japan, pp. 185–189 (2014). <https://doi.org/10.1109/IIAI-AAI.2014.47>
20. Sassi, I.B., Yahia, S.B., Liiv, I.: MORec: at the crossroads of context-aware and multi-criteria decision making for online music recommendation. *Expert Syst. Appl.* **183**, 115375 (2021). ISSN: 0957-4174. <https://doi.org/10.1016/j.eswa.2021.115375>

21. Lindawati, Chen, W., Cui, W., Hari, N.: Feasibility analysis on collaborative platform for delivery fulfillment in smart city. In: Proceedings of the 2015 IEEE International Conference on Smart City/SocialCom/SustainCom, pp. 147–152. IEEE (2015)
22. Wang, R., et al.: TaxiRec: recommending road clusters to taxi drivers using ranking-based extreme learning machines. *IEEE Trans. Knowl. Data Eng.* **30**(3), 585–598 (2018)
23. Comi, A., Rosaci, D.: SMARTSAN: a P2P social agent network for generating recommendations in a smart city environment. In: Proceedings of the 18th Workshop “From Objects to Agents”, CEUR Workshop Proceedings, pp. 108–112 (2017)
24. Kolstad, A., Özgöbek, Ö., Gulla, J.A., Litlehamar, S.: Content-based recommendations for sustainable wardrobes using linked open data. *Mob. Netw. Appl.* **2018**, 1–8 (2018)
25. Horng, G.J.: The adaptive recommendation segment mechanism to reduce traffic congestion in smart city. In: Proceedings of the 10th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 155–158. IEEE (2014)
26. Çano, E., Morisio, M.: Hybrid recommender systems: a systematic literature review, 1487–1524 (2017)
27. Handte, M., Foell, S., Wagner, S., Kortuem, G., Marrón, P.J.: An internet-of things enabled connected navigation system for urban bus riders. *IEEE Internet Things J.* **3**(5), 735–744 (2016)
28. Toader, B., Moawad, A., Fouquet, F., Hartmann, T., Popescu, M., Viti, F.: A new modelling framework over temporal graphs for collaborative mobility recommendation systems. In: Proceedings of the 20th IEEE International Conference on Intelligent Transportation Systems, pp. 1–6. IEEE (2017)
29. De Meo, P., Quattrone, G., Terracina, G., Ursino, D.: A multi-agent system for the management of e-government services. In: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, pp. 718–724 (2005)
30. Terán, L.F., Meier, A.: A fuzzy recommender system for elections. In: Proceedings of the 1st International Conference on Electronic Government and the Information Systems Perspective, pp. 62–76 (2010)
31. Lu, J., Shambour, Q., Zhang, G.: Recommendation technique-based government-to-business personalized e-services. In: Proceedings of the 2009 Annual Conference of the North American Fuzzy Information Processing Society, pp. 1–6 (2009)
32. Gómez-Martínez, E., et al.: A semantic approach for designing assistive software recommender systems. *J. Syst. Softw.* **104**, 166–178 (2015)
33. Lemlouma, T., Chalouf, M.A.: Smart media services through TV sets for elderly and dependent persons. In: Godara, B., Nikita, K.S. (eds.) *MobiHealth 2012*. LNICTS, vol. 61, pp. 30–40. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37893-5_4
34. Kinawy, S., El-Diraby, T., Konomi, H.: Customizing information delivery to project stakeholders in the smart city. *Sustain. Cities Soc.* **38**, 286–300 (2018)
35. Kavanaugh, A., Krishnan, S., Pérez-Quinones, M., Tedesco, J., Madondo, K., Ahuja, A.: Encouraging civic participation through local news aggregation. *Inf. Polity* **19**(1, 2), 35–56 (2014)
36. Madani, R., Ez-zahout, A., Idrissi, A.: An overview of recommender systems in the context of smart cities. In: 2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech), pp. 1–9 (2020). <https://doi.org/10.1109/CloudTech49835.2020.9365877>



Sentiment Analysis for Competence-Based e-Assessment Using Machine Learning and Lexicon Approach

Mohammed Amraouy¹(✉), Mostafa Bellafkih¹, Abdellah Bennane²,
and Jallal Talaghzi¹

¹ National Institute of Posts and Telecommunications, Rabat, Morocco
amraouy.mohammed@inpt.ac.ma

² Inspectors Training Center for Education (CFIE), Mohammed V University, Rabat, Morocco

Abstract. Most education systems adopt the competence approach as a reference framework for learning and teaching processes. In consequence, learning outcomes should be evaluated in terms of competencies development. In this paradigm, e-assessment process must take into consideration the different dimensions of competence, especially the cognitive and socio-affective dimensions. This paper emphasizes the importance of machine learning approaches and lexicon-based approach to detect the socio-affective component, based on sentiment analysis of learners' interaction messages. For classification and comparative analysis, TextBlob and Naïve Bayes algorithms have been used. Based on the information gathered, the results of this study indicate that it contributes significantly to providing a systemic approach to evaluate learning outcomes.

Keywords: Sentiment analysis · Lexicon approach · Machine learning · E-assessment · Competence · Digital learning

1 Introduction

Due to evolution of internet and web technology, E-learning platform and social media has become the place for online learning and exchanging ideas [1]. In an online learning environment, teachers use discussion forums as a tool for communication, collaboration and sharing to meet the requirement of active and digital pedagogy. Each message gives an idea of the behavioral, cognitive and affective learner's engagement. The competence-based approach as a theoretical framework for active learning requires that the learning assessment process must take into consideration both the cognitive and socio-affective component of the competence to be assessed [2]. The first component is verifiable by the conventional tools (QSM, Project, etc.) [3]. On the other hand, socio-affective component evaluation invites teachers to analyze learners' digital traces [1]. However, in front of a large number of learners it will be difficult and even impossible for the teacher to manually analyze the large amount of data. "*Communities of learning face two major problems for their development, low quality and quantity of participation and information overload due to high amount of participation*". A possible solution is to use

text mining techniques in general and sentiment analysis in particular to detect affective engagement [1, 4, 5].

Sentiment analysis is imperative and essential to our technological development for the benefits that it can provide (social media sentiment analysis, teaching and learning sentiment analysis, movie reviews analysis, news sentiment analysis, etc.). There are many related names of sentiment analysis such as opinion analysis, sentiment mining, subjectivity analysis, affect analysis or emotion analysis. In this work, we use sentiment analysis to detect learners' affective engagement, based on their interaction messages collected from two different online courses.

There are three main approaches that can be implemented in sentiment analysis which are machine learning approach, hybrid approach and lexicon-based approach [6]. This research uses TextBlob approach for lexicon and Naïve Bayes approach for machine learning, in order to identify learners' affective engagement, by classifying messages interaction into three classes: positive, negative and neutral.

The objectives of this work are: first one is to apply text sentiment using TextBlob approach, second one is to construct a classifier model of sentiment analysis using Naïve Bayes approach, the third one is to evaluate the performance of both approaches and the last one is to integrate these two methods into an E-learning platform in order to guarantee a systemic evaluation consistent with the competence-based approach, in the sense that it takes into consideration the socio-affective component in the online learning assessment process.

The remainder of this paper is organized as follows. Section 2 presents the background to the related literature of sentiment analysis. Section 3 presents our structure, steps and method. Results and discussions are presented in Sect. 4. Finally, Sect. 5 highlights the conclusion and discusses future studies.

2 Background

Currently the applications of sentiment analysis are mostly commercial “The current reality is that AI innovations are not yet adequately directed at the common good of humanity. For example, in a 2021 survey circulated to all Member States by UNESCO on the use of AI to support learning continuity and quality during COVID-19, only five countries reported that they had used AI tools in their education response to the crisis (UNESCO, 2021b). While there are increasing numbers of commercial AI tools being developed and implemented across institutions, there are very few AI tools or practices orientated to the common good of education and humanity” [7]. Movie producers use sentiment analysis to evaluate audience feedback on their projects. Companies include AI technology to assess how consumers are reacting to their products and services. In education context, sentiment analysis could be used to facilitate the process of competencies assessment in an online learning situation. Indeed, learning evaluation in general and competence assessment in particular, remains a major and complex problem [3].

In a competence-based approach, e-assessment process must take into consideration the different dimensions of competence, especially the cognitive, socio-affective and technological dimensions. Figure 1 gives an illustration the different elements and components of the competence based e-Assessment proposal.

Each learner has a competencies profile made up of the competencies already validated. This competencies profile is part of the learner's model. It contains the following declarative knowledge:

- *idc*: id of the competence to be evaluate;
- *cvl*: Competence validation level;
- *cld*: Competence level development

$$cld = \frac{(w_1 * alcc + w_2 dlsc)}{w_1 + w_2}$$

- *alcc*: acquisition level of the cognitive component;
- *dlsc*: development level of the socio-affective component;
- *w*: Weight of each component in relation to competence development.

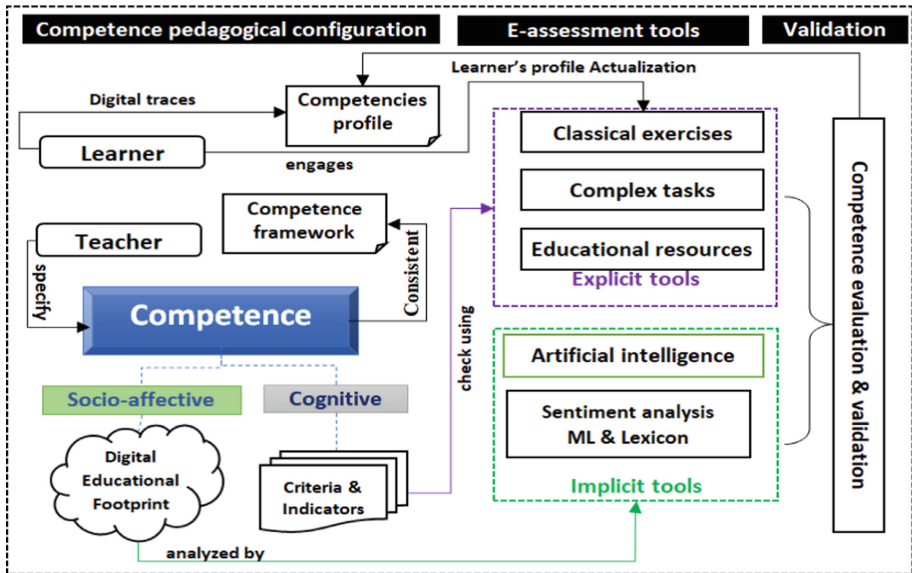


Fig. 1. Competence based e-Assessment proposal

Teacher selects or defines the competence to be assessed, while specifying its cognitive and socio-affective dimension. The weighting of each of these two dimensions is determined by the teacher or by the training manager. Any learning path in a competence-based approach can be represented as: (*idc*0, *cvl*0, *cld*0), (*idc*1, *cvl*1, *cld*1), ..., (*idc**n*-1, *cvl**n*-1, *cld**n*-1).

To assess the cognitive component, criteria and indicators are mobilized to specify the learning outcomes. They will be checked using explicit tools: Classic exercises, complex tasks (Open-ended question, Project or/and Portfolio) [3].

The socio-affective component is evaluated through sentiment analysis of the digital traces (the resources consulted, the number, the duration of collaborative interactions, results obtained) of the learner via the mobilization of artificial intelligence techniques [4]. In this sense, Halimi and H. Seridi-Bouchelaghem indicate that competence oriented e-Assessment requires the mobilization of sophisticated technological tools « *We believe that this type of assessment requires deep changes in the design of online evaluation systems through the use of recent and sophisticated computing technologies, namely: semantic web, ontologies, learning analytics approaches, artificial intelligence, data mining and data visualization tools, these technologies offer a promising way to assess students' competencies* » [8].

3 Research Methodology

In online learning situation, Sentiment Analysis mobilizes machine learning approach and/or lexicon based approach to identify and classify the opinions as expressed in a text, comments, chats or question about a course or any learning task in order to determine if the opinions or sentiments are positive, negative or neutral. In this research, we use TextBlob for lexicon approach and Naïve Bayes approach for supervised machine learning approach. Figure 2 illustrates the methodology used.

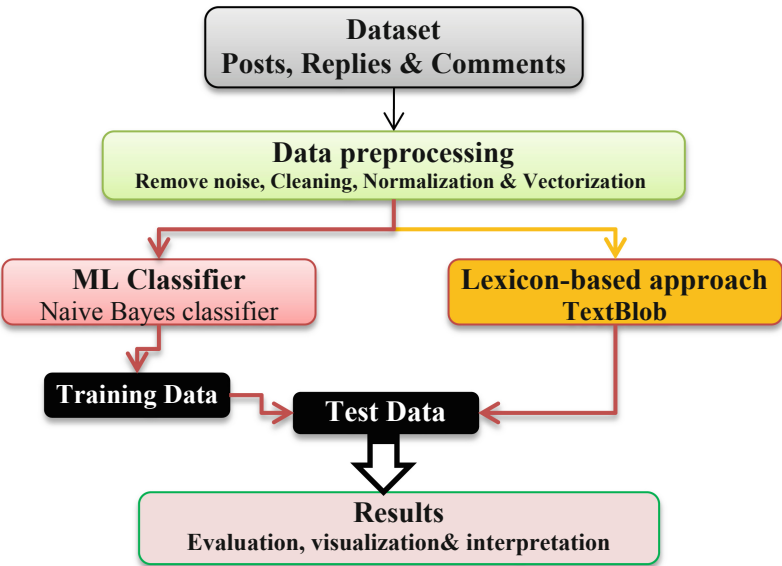


Fig. 2. Sentiment classification methodology

3.1 Overview of Data

In this research, we used Posts dataset from two online courses. The first one is about web development and the second is about digital learning paradigm, both are carried out at

Faculty of Letters and Human Sciences Oujda. Dataset contains the messages extracted from February 01 until June 30 in 2022. The data collected in this test represents a database of 246 samples that contains lots of questions, answers, opinions, comments shared by learners in different ways. The data-set size augmented to 2216 interactions messages from the 246 real messages; the rest is generated based on a data augmentation with Back Translation technique. Then, we split data to data testing (30% of the complete data set) and data training (70% of the data).

Figure 3 shows the most used words. These words generalize the topics in messages discussion such as learners are talking about the shared videos, work requested by the teacher, thanks, LMS, Learning, Classroom, and Technology.

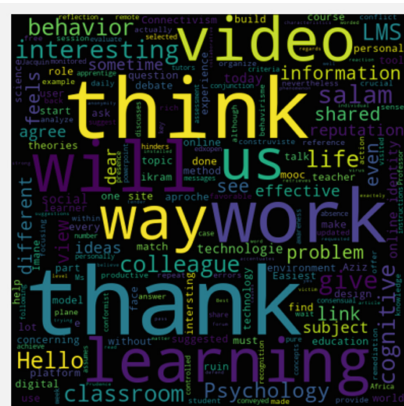


Fig. 3. Word-cloud for dataset

3.2 Data Preprocessing

Data preprocessing is a crucial phase in the data mining process. In this work, we used four transformations methods: Noise removal, Text cleaning, Normalization and Vectorization text.

- Noise removal and text cleaning include removing special characters, digits and unnecessary information like links, tags and emoticons.
- Text normalization is the process of transforming the text into a consistent form, we use two main techniques: Stemming and lemmatization. In the data collected, learners usually begin their messages with “salam”, “salam3alayom”, “hello, salam”. These words will be replaced by “salam” or simply “hello”
- Text vectorization: turning the text into a numerical representation. Like the other operations, Vectorization is taken care of automatically using NLP methods.

After that, we have classified learners’ messages into three classes: positive (text indicates joy, excitement, kindness, thanks), Negative (text indicates sorrow, aggression,

discrimination) and Neutral (text expresses no personal views and merely transmits information)

3.3 Data Classification

Sentiment detection of learners' interaction messages has been performed using TextBlob for lexicon-based approach and Naive Bayes algorithm for machine learning approach.

TextBlob is a Python-based NLP library that uses a natural language toolkit (NLTK). It can be used for complex analysis on textual data (language translation, speech tagging, tokenization, sentiment analysis, etc.) [9]. TextBlob supports several languages including Arabic and French; it is among the reasons that push us to choose this tool. The sentiment function of TextBlob returns tow outputs, which are polarity and subjectivity. Polarity is a float between $[-1, 1]$ where 1 refers to positive sentiment and -1 means a negative sentiment. Subjective is, also, a float that lies within $[0, 1]$ and refers to personal opinion or judgment. Lower subjectivity means that the text contains factual information. In this phase we have classified leaners interaction messages into three classes: positive, neutral and negative, based on polarity and subjectivity.

Naive Bayes classifier is a supervised machine learning technique that works on the principals of conditional probability as given by the Bayes' theorem as formulated by Eq. 1.

$$P(A \setminus B) = \frac{P(B \setminus A)P(A)}{P(B)} \quad (1)$$

This theorem gives the conditional probability of an event A given another event B has occurred. In text classification each document/sentence represents by vector of words, in order to calculate its probability using the Eq. 2. This equation is used by Naive Bayes classifier in the training phase.

$$P(doc \setminus V_j) = \prod_{i=1}^{length(doc)} P(a_i = w_k \setminus v_j) \quad (2)$$

where $P(a_i = w_k \setminus v_j)$ is the probability that a word in position i is w_k ; one more assumption is that: $P(a_i = w_k \setminus v_j) = P(a_m = w_k \setminus v_j) \forall i, m$

The training phase allows the classifier to determine the probabilities for all possible words in the dataset vocabulary. After that, in the testing phase Naïve Bayes classifies a new document/sentence according to Eq. 3 [10].

$$V_{NB} = argmax_{v_j \in V} P(v_j) \prod_{w \in words} (P(w \setminus v_j)) \quad (3)$$

where V stands for value or class

4 Results and Analysis

A comparative analysis is carried out using Naïve Bayes approach for machine learning and TextBlob for lexicon approach. Figure 4 shows pie cart of all TextBlob annotated 246 messages after classification, showing 52% positive messages, 35% neutral and 13% negative sentiment. For the same data, human experts assign 65% neutral messages, 29% positive messages and 6% negative messages (see Fig. 5).

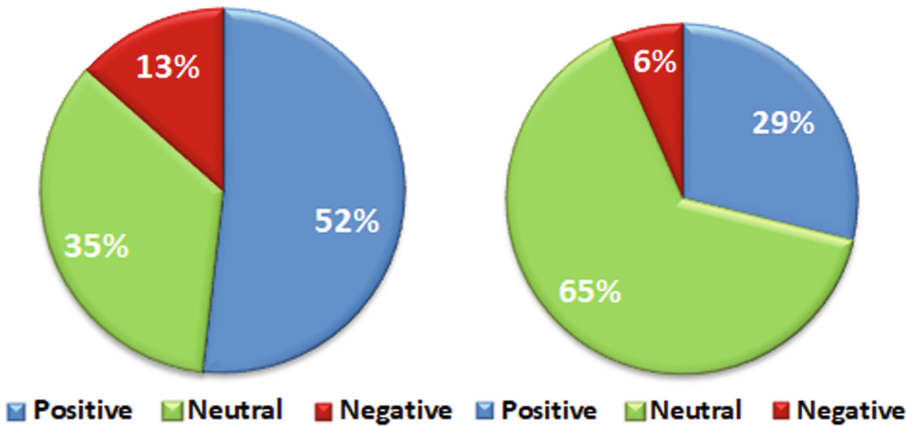


Fig. 4. Pie chart of TextBlob annotation

Fig. 5. Pie chart of manual annotation

The difference between TextBlob and manual classification is because data contains a lot of factual information about digital learning and web development. For example, learners use words like “very” “excellent” to express the quality of a response or resource (link, video, and document). For human expert will be a neutral message, on the other hand for TextBlob will be a positive message, due to the use of the adjective. The table shows some examples (Table 1).

After that we also applied the Naïve Bayes sentiment classifier to the corpus. The graphs in Figs 6, 7 and 8 shows the performance of the two approaches used in this work based on Precision, Recall ad F1-score [11].

Considering these three parameters, The Naïve Bayesian classifier had highest performance with precision (52% for negative class, 90% for Neutral class and 100% for Positive), recall (77% for negative class, 91% for Neutral class and 83% for Positive), and F1-score (62% for negative class, 91% for Neutral class and 91% for Positive) compared to TextBlob tool. The result cannot be considered sufficient for now to confirm the maturity of the Naïve Bayesian approach, because of the data set size. We started with a tiny dataset composed from only 246 learners’ interaction messages, to overcome this limitation we have amplified the original to 2216 using a data-augmentation technique. Despite this, it remains relatively low for model validation purpose.

To give more insight into the quality of the results, we realized a PHP web application where we integrated TextBlob classifier and Naïve Bayes pre-trained model into our E-Dirrassa learning content management system (LCMS) [3]. TextBlob used for language detection and translation and Naïve Bayes classifier for sentiment analysis detection.

Table 1. Examples of interaction messages and their classification

Interaction messages	Type of classification	
	Manual	TextBlob
The video “The Easiest way to ruin your reputation online” explains how build social media design	Neutral	Positive (0.03,0.06)
The ability to draw distinctions between important and unimportant information is vital	Neutral	Positive (0.03,0.08)
ELearning also called online learning refers to the total or partial use of some type of information network to deliver course content	Neutral	Negative (−0.05,0.5)
You are right Miss Reitha it’s difficult to ...	Positive	Negative (0.11,0.77)
I consulted a few articles about reinforcement and punishment	Neutral	Negative (0.9,0.35)

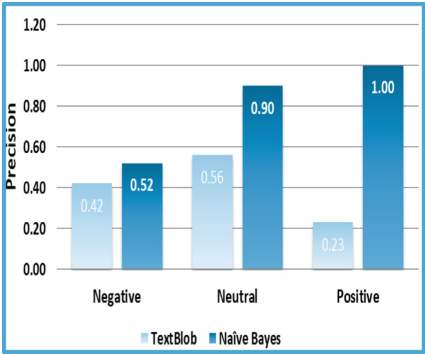


Fig. 6. Precision of TextBlob and Naïve Bayes

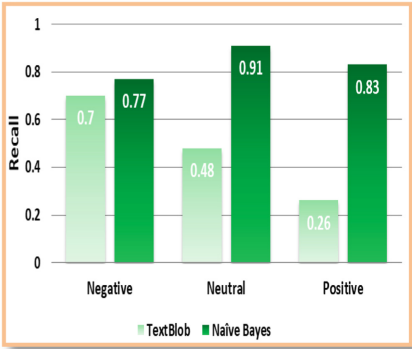


Fig. 7. Recall of TextBlob and Naïve Bayes

Figure 9 shows sentiment analysis of learner messages; we extracted the message from the database related to his owner. Thus, showing the message, polarity, subjectivity and sentiment, will give a chance for the tutor to validate and give more judgment of each message. The sentiment analysis allowed us to create a dashboard concerning the different components of the competence. This dashboard displays a dynamic analysis concerning the different activities carried out by the learner. This meets the requirements of the online learning assessment model proposed in this study.

The analysis of sentiment in online learning will shed light on the learner’s affective engagement as a compound of competence based e-assessment process. Also, this will give the opportunity to the tutor and the system to track the sentiment satisfaction, orientation of each subscribed learner in the course and also to create adaptive learning or assessment process [12, 13].

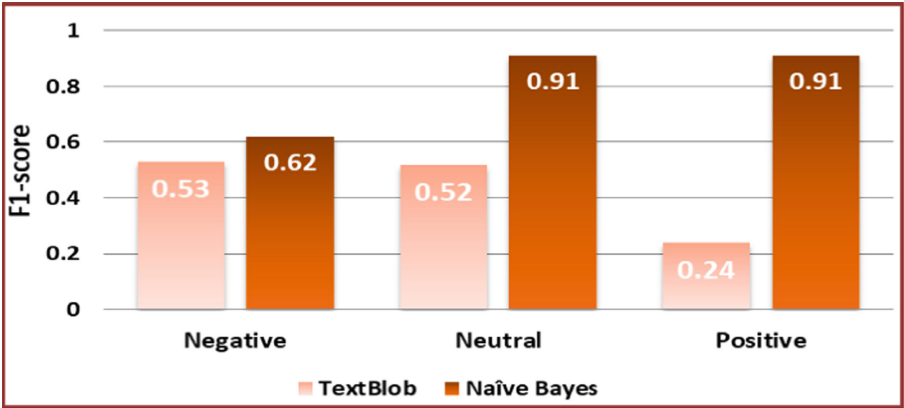


Fig. 8. F1-score of TextBlob and Naïve Baye

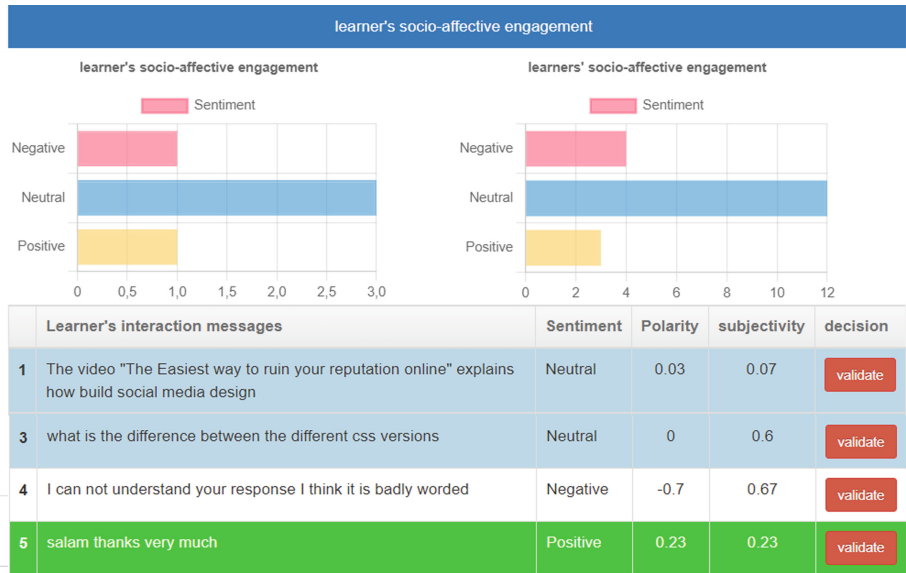


Fig. 9. Sentiment analysis of a learner's messages

5 Conclusions and Future Work

In online learning situation, Sentiment Analysis mobilizes machine learning approach and/or lexicon based approach to identify and classify the opinions as expressed in a text, comments, chats or question about a course or any learning task in order to determine if the opinions or sentiments are positive, negative or neutral. In this research, we conducted a comparative study between Naïve base classifier and TextBlob tools to build a model able to detect the learner's affective engagement. The results presented in this paper show the performance of the machine learning model for text sentiment

analysis. In fact, multinomial Naïve Bayes classifier algorithm performs better with accuracy 87%. Therefore, this model will present an efficient tool for the tutor and the system to gauge the learner's affective engagement via the posted messages throughout the learning process.

This research also allowed us to integrate sentiment analysis tools into our E-Dirrassa LCMS platform [3], in order to promote an online evaluation process consistent with the competence based approach. In future work, we will use sentiment analysis to create an adaptive framework for Competence based e-Assessment.

References

1. Amraouy, M., Bennane, A., Himmi, M.M., Bellafkih, M., BenOmar, A.: Detecting the learner's motivational state in online learning situation towards adaptive learning environments. In Proceedings of The 13th International Conference On Intelligent Systems: Theories and Applications, pp. 1–6. Article 22 (2020). <https://doi.org/10.1145/3419604.3419760>
2. Beatriz, E.F.G., Baldiris, S.M., Fabregat, R.: A set of software tools to build an author assessment package on Moodle implementing the AEEA proposal. In: 10th IEEE International Conference on Advanced Learning Technologies, pp. 2–4 (2010). <https://doi.org/10.1109/ICALT.2010.26>
3. Amraouy, M., Bellafkih, M., Bennane, A., Himmi, M.M.: Online competency-based assessment (OCBA): from conceptual model to operational authoring system. *Int. J. Interact. Mob. Technol.* **16**(04), 46–57 (2022). <https://doi.org/10.3991/ijim.v16i04.28373>
4. Wahyono, I.D., Asfani, K., Mohamad, M.M., Saryono, D., Putranto, H., Said, M.N.H.B.M.: Text mining in chat room of online learning for detection emotion using artificial intelligence. In: International Conference on Computer Science, Information Technology, and Electrical Engineering, ICOMITEE, pp. 63–67 (2021). <https://doi.org/10.1109/ICOMITEE53461.2021.9650099>
5. Kelley, P., Gavrin, A., Lindell, R.S.: Text mining online discussions in an introductory physics course. In: Physics Education Research, 2018, no. March, pp. 216–219 (2018). <https://doi.org/10.1119/perc.2017.pr.049>
6. Birjali, M., Kasri, M., Beni-hssane, A.: A comprehensive survey on sentiment analysis : approaches, challenges and trends. *Knowl. Based Syst.* **226**, 107134 (2021). <https://doi.org/10.1016/j.knosys.2021.107134>
7. Miao, F., Holmes, W.: International Forum on AI and Education: Ensuring AI as a Common Good to Transform Education (2022)
8. Kumar, V., Troussas, C. (eds.): ITS 2020. LNCS, vol. 12149. Springer, Cham (2020). <https://doi.org/10.1007/978-3-030-49663-0>
9. Umair, M., Hakim, A., Hussain, A.: Sentiment analysis of students feedback before and after COVID-19 Pandemic. *Int. J. Emerg. Technol.* **12**(2), 177–182 (2021)
10. Adesh, N.D.: Sentimental analysis of student feedback using machine learning techniques. *Int. J. Recent Technol. Eng.* **8**, 986–991 (2019)
11. Amra, I.A.A., Maghari, A.Y.: Students performance prediction using KNN and Naïve Bayesian. In: ICIT 2017 8th International Conference on Information Technology Proceedings no. December 2018, pp. 909–913 (2017). <https://doi.org/10.1109/ICITECH.2017.8079967>
12. Chanaa, A., El Faddouli, N.: E-learning text sentiment classification using hierarchical attention network (HAN). *Int. J. Emerg. Technol. Learn.* **16**(13), 157–167 (2021)
13. Talaghzi, J., Bennane, A., Himmi, M.M., Bellafkih, M., Benomar, A.: Online adaptive learning: A review of literature. In: ACM International Conference Proceeding Series, no. December, pp. 115–120 (2020). <https://doi.org/10.1145/3419604.3419759>



User Sentiment Analysis Towards Adapting Smart Cities in Egypt

Lamiaa Mostafa^(✉) and Sara Beshir

Business Information System Department, Arab Academy for Science and Technology and
Maritime Transport, Alexandria, Egypt
Lamiaa.mostafa31@aast.edu

Abstract. The smart city combines physical infrastructure to create mobility, add efficiencies, save energy, enhance air and water quality, recover from disasters, fix problems, data collection, data sharing, and effective resource deployment. Citizens are motivated in Egypt to be smart city's residents, and the government should work on its citizens' satisfaction and fulfill their needs when constructing and developing smart cities. This research aimed to investigate the factors that affect citizens choosing smart city life. The Sentiments were collected from 7,000 residents in 5 settlement compounds in New Cairo. Sentiments were collected, analyzed, and categorized into the following 10 groups: reducing traffic congestion, increasing people's safety, reducing air pollution, reducing travel costs, advanced technology, parking availability, technical Facilities, Green landscape, reducing Noise pollution, and positive energy and relaxation. The results showed a positive attitude towards applying smart cities in Egypt. NB has the highest accuracy, 98%, SVM, 88%, and the decision tree, 71%.

Keywords: Smart city · Government · Sentiment · Citizen · Egypt

1 Introduction

The smart city not only applies technology; it combines people and technology to solve problems in other aspects, such as social, economic, and sustainability. In smart cities, data must be collected and shared continuously since the technologies are changing fast. Governments who plan to adopt smart cities must understand the requirements of high-quality social interactions and social relationships that assist smart cities in connecting with their residents and the outside world. The government is seeking smart cities to solve the problem of city problems. Smart Cities' main properties are social and economic activities. In 2050, 68% of the world's population will live in these cities [1].

The main problems of Urban places are energy consumption, pollution, and complex waste management, that Smart cities can solve. Different Countries built Smart cities, including Spain [4], Brazil [2], Ireland [5], South Korea [3], and Japan [6].

A smart city combines physical infrastructure with social, economic, and sustainable factors; all these factors will be used by the "Smart people" who are citizens with active participation and engagement in various communities [7, 8]. Smart people allow sharing

of information and social interaction that makes the leaders and people of the city be alert and responsive during emergencies. Smart city owners develop internet websites and mobile applications for their resident's communication [9].

Continuous interactions between government citizens are one of the main smart city success factors. The effective development of a smart city needs citizen participation socially and politically [10]. The importance of citizen participation can affect the effectiveness of smart city construction [11, 12].

The rapid development of social media, such as Twitter, Facebook, and Instagram, makes different users feel free to express their opinions and emotions. It considers an effective path for participation in smart city construction [13]. Analyzing sentiments can help smart city construction team to build a city with the requirements of the users [14].

This research collected 7,000 sentiments from smart residents in 5 settlement compounds in New Cairo. The research results will help understand the Egyptian citizen public's cognitive status about living in a smart city. The conclusions can assist smart city construction teams in building a smart city with the same features required from their future residents. The novelty and contributions of this search are mainly in the following two aspects. First, create a model to collect user sentiments and analyze them using text mining methods. Second, understanding the perspective of Egyptian citizens on their future smart city needs, Ranking the factors that affect the Egyptian resident to live in the smart city. Egyptian government can further understand Egyptian people's requirements while constructing the smart city.

The paper is organized as follows: Sect. 2 provides a literature review including background about smart cities in the world, the Internet of Things, Smart Cities in Egypt, and Sentiment analysis. Section 3 provides details about the Sentiment Analysis Model and its components. Section 4 conducts the experimental results, and Sect. 5 discusses the conclusions, limitations, and future directions.

2 Literature Review

Smart cities have drawn a lot of attention over the past few years. This is a result of the fact that smart cities use several smart systems to help improve this urban-development project's roads. It uses the Internet of Things (IoT), which has grown to be a great topic for businesses and researchers. The city's quality of life and road infrastructure has been improved with the aid of several IoT-enabling technologies, like Wireless Sensor Networks (WSN). In this section, several IoT and smart cities concepts will be discussed.

2.1 Smart Cities

Pilot smart city projects have started to be developed and marketed internationally in recent years [15]. The concept of "smart cities" has gained popularity over the past few years, with its major goal being the creation of protocols and technology that would enable us to create cities that are more responsive, dynamic, and engaging to enhance daily living. To realize the whole goal of smartness in modern cities, infrastructure, transportation, utilities, and many other services must be greatly upgraded or rebuilt. It is important to emphasize that the development of smart cities requires not only the

building of high-tech infrastructure and services but also the involvement of citizens in the provision and improvement of services [16].

Smart city advantages come with new potential risks. IoT devices largely reliant on important infrastructures, such as water networks that might affect the health and safety of populations, are vulnerable to cyber attacks [17]. To reduce the risk of assaults on information and operational technologies, it is important to monitor and manage utility networks continuously. By doing so, cities will become smarter and more effective living places [18, 19].

According to [20] and [21], for citizens to adopt smart city services, they must believe they can use them. Prior experience with technology viewed as comparable to the Smart City is a significant element.

2.2 Internet of Things (IoT)

IoT is one of the most endearing technologies. Utilizing any links and services enables people and things to be connected to anyone and everything, anywhere, at any time. To deliver developed services, it helps to give a space for sensors and devices to be integrated easily into a smart environment, as well as intelligent services for people [22].

Smart utilities have greatly benefited from recent developments in the Internet of Things, big data analytics, cloud computing, social media, and smart meters. IoT is steadily turning into a crucial part of communication technology. Smart nodes, which include devices, sensors, services, and applications, can interact and communicate in real time more and more with the help of IoT, making them a crucial part of Industry 4.0. [23].

The ecosystems for smart utilities on IoT-based contain new participants whose primary identification is digital; as a result, stakeholders face several difficulties regarding the security of data, infrastructure, and services. Consequently, key concerns for smart utilities include security, confidentiality, and authentication [24].

Operators should use continuous evaluation and performance prediction to maintain the high efficiency of IoT systems. Given that the IoT is extremely complex due to the vast number of connected devices, the enormous size of exchanged data, high-speed network topology exchange, and heterogeneous devices and connected things, the prediction of IoT system performance is regarded as a significant challenge [25].

IoT has significantly reduced the human labor required in practically every sector. The Internet will be connected to trillions of machine-type devices, such as connected cars, wearable sensors, or everyday objects, creating a vast IoT ecosystem. IoT makes it possible for gadgets to communicate via wireless networks, gather data, and process it instantly. Smart homes, smart grids, smart agriculture, smart healthcare, smart streets, and parking are examples of IoT applications in smart cities. IoT still has a long way to go before it can fully realize applications for smart cities. However, due to issues with computing, transmission, local real-time data analysis, end-to-end (E2E) latency, huge device connectivity, and privacy. Additionally, IoT devices have a finite number of resources and energy available to them [26].

IoT technology, which enables the communication between passive and active devices or items used for daily purposes in smart cities, is a key component of the

general vision of sustainable smart cities. This technology enables services to be provided to city residents in a sustainable and balanced manner. In all interdisciplinary fields, the theory and vision of future sustainable, intelligent cities are intimately tied to IoT technology and communication networks. This is because IoT communication networks have a substantial impact on how sustainable smart cities are applied, influencing their scalability, stability, and accuracy of computational output [27].

2.3 Smart Cities in Egypt

Egypt is a growing nation, and several environmental, social, and economic problems plague its cities. The Egyptian government faces these obstacles to attain sustainability and raise the quality of life for citizens in the nation's vision for the future metropolis. Egyptian stakeholders were motivated to focus on the sustainable smart city concept for the next generation of Egyptian towns by the fourth cyber-physical revolution and the exponential growth in ICT advances. Smart cities can help sustainability through their numerous potentials. Given the current conditions, Egypt's adoption of smart city trends is essential [28].

According to [29], the research aims to establish which smartness efforts designate Egyptian towns as "Smart Cities. The 40 experts give the weighted average rating for each of the twenty-two components of smart systems in smart Egyptian cities. In all smart cities literature, twenty-two major smart system components were listed in the interview. To discover the key traits of the smart Egyptian cities, the twenty-two provided smart component systems were categorized according to the key smart initiatives of smart cities, and the average scoring of each smart initiative was calculated. The findings indicate that the key initiatives of smart Egyptian cities are smart governance, smart mobility, and infrastructure.

One of the most obvious goals of these future cities is to improve citizens' quality of life while also helping to address the expanding population issue while meeting people's basic requirements [30].

Providing people with high-quality access to electricity, water, food, transportation, education, healthcare, and governmental services, as well as preserving the environment's status and achieving sustainability, are the Egyptian government's two key goals for future cities [31]. The government is confident that these goals would be met by implementing the new smart city idea for Egypt's fourth-generation cities [32]. To address these issues, Egypt's vision for 2052 calls for the construction of new sustainable cities with a total area of 580,000 acres that can house approximately 30 million people, all while utilizing smart technology to improve the quality of life of its residents by providing the bare minimum of essential infrastructural facilities with better environmental conditions and better quality of life, as well as supporting health and socio-cultural infrastructure. The new cities aspire for sustainability by properly managing their natural resources and utilizing them and leveraging technology to address everyday problems and create a higher standard of living for all future Egyptian city generations.

On behalf of the Egyptian government, the New Urban Communities Authority (NUCA) has been attempting to relocate some of the capital's 20+ million residents to twenty new, fourth-generation smart towns scattered throughout Egypt. They include the New Administrative, new 6th of October City, an extension of Sheikh Zayed, an

extension of New Nubaria in Behaira Governorate; new Bir al-Abd in northern Sinai; new Al-Fashn in BeniSuef; new AlWarraq in Giza; New Alamein city in Marsa Matrouh; new Mallawi in Minya Governorate; new Mansoura City in Dakahlia Governorate; East Port Said “Salam” in Western Port Said; October Gardens, El-QalyubiaGovernorate; Western Assiut; El-Obour City, Western Qena; New Luxor; Capital and South of New Cairo in Cairo, New Sphinx, and Toshka in Aswan [33], Table 1 summarizes the citizen benefits of smart cities.

Table 1. Citizen benefits of smart cities

Literature review	Previous researchers
Reducing traffic congestion	18,19
Increasing people safety	19,20,21,29
Clean Place	16, 17,18,19
Reduce travel cost	19, 21,22
Advanced technology	19, 10,15,16,17,18
Parking availability	10,16,19
Technical facilities	10, 17, 18, 19, 20,21
Green landscape	10, 19,20,21
Reduce noise pollution	10,19, 20,21
Positive energy and relaxation	10,18,19

2.4 Sentiment Analysis

Sentiment analysis enables you to identify the broad feelings that go along with a specific situation without having a deeper understanding of the underlying causes [34].

According to [35], sentiment analysis of four smart city initiatives’ positive feelings is constantly and strongly expressed about the smart city project. These results are in line with what has been noticed in other empirical investigations. For instance, people in Greece and Cyprus who were unfamiliar with local smart city applications (like situations in Africa) reported having a keen curiosity about how a smart city may enhance their quality of life [36, 37].

Researchers in [10] used the CNN-BiLSTM model, which was previously used to accumulate the emotional value of each comment and calculate the average value to obtain the public’s perception of “smart city” in each year to compare the evolution trend of various users types on the microblogging platform for the emotional polarity of smart cities. Additionally, the study divided the obtained samples into Weibo official certified users, and non-Weibo official certified users according to user types, namely Weibo masters, Weibo members, and other types, to compare the emotional tendencies of various users on the topic of “smart cities.

Sentiment analysis passes through three main stages: pre-processing, feature selection, and classification. Pre-processing removes the unneeded words, features selection method chooses the most important keywords, such as Term Inverse Term Frequency (TFIDF) and Bag of words (BOW). The classification divides sentiments into two classes in the training stage and then classifies new sentiments into one of the two classes. Three main classifiers are usually used in sentiment classification Support Vector Machine (SVM) [10, 38], NB classifier (NB) [39, 40], and Decision Tree classifier [41, 42]. Naïve Bayes text classification model assumes that all attributes are independent of each other given the context of the class, Support vector machines select the best possible surface to separate positive and negative training samples. Decision Tree examines categories with the same characteristics.

3 Citizen Sentiment Analysis Model

The Citizen Sentiment Analysis Model comprises three stages: data collection, processing, and classification. Sentiments are collected from 7000 Egyptian residents of 6 settlement compounds. The sentiment model classified the sentiments into one of the 10 categories of the benefits of living in a Smart city. The sentiment model classifies each sentiment into one of these ten groups: reducing traffic congestion, increasing people’s safety, Reducing air pollution, reducing travel costs, advanced technology, parking availability, technical Facilities, green Landscape, reducing Noise pollution, and positive energy and relaxation. Table 2 lists the keywords exceeding 500 times in the 7000 sentiments mapped to the smart city benefits category.

Table 2. List of keywords collected from Sentiments that exceed 500 times for each category

Category	Keyword 1	Keyword 2	Keyword 3	Keyword 4	Keyword 5
Reducing traffic congestion	Traffic	Busy	deadlock	Congest	Car
Increasing people safety	Safe	Secure	more	Infect	Parent
Clean place	Minimize	Harm	Shrink	Air	Pollution
Reduce travel cost	Travel	Cost	Expense	Location	Move
Advanced technology	Science	Develop	Technology	Digital	Internet
Parking availability	Park	Space	Car	Free	Available
Technical facilities	Technical	Equipment	Facility	Service	Technique
Green landscape	Layout	Green	Park	Plant	Tree
Reduce noise pollution	Noise	Harm	Sound	High sound	Quiet
Positive energy and relaxation	Positive	Energy	Calm	Relax	Vibe

The processing step is designed using Knime [43]. Knime is an analytical tool used in web mining [42, 43]. The sentiment model is divided into three steps: processing, feature selection, and classification, as shown in Table 3; each component is described in the following paragraphs.

Table 3. Citizen sentiment model stages

Data collection
Processing (Porter Stemmer)
Feature Selection (TFIDF)
Classifiers (SVM-NB-Decision Tree)

Pre-processing clean sentiments and removing the stop words, feature selection defines the highest occurrence keywords and the classification that collects items with the same properties. BOW [41] and TF-IDF [39] are used for the features selection method in the citizen sentiment model. Classifiers used are SVM, NB, and Decision Tree Classifiers will be validated using the following measures: accuracy, precision, and recall had been used [34, 35].

4 Experimental Results

Classifiers are validated using Accuracy, Precision and Recall, represented on the following equations [38, 41]:

$$\text{recall} = \frac{\text{number of relevant items retrieved}}{\text{a number of relevant items in the collection}}$$

(1)

$$\text{precion} = \frac{\text{number of relevant items retrieved}}{\text{total number of items retrieved}}$$

(2)

Citizen reviews were classified into two classes: Resident and not resident. The sentiment analysis classifier tests the machine learning classifiers based on 7000 sentiments; sentiments are classified into resident sentiment and no resident sentiment. NB, SVM, and Decision Tree were used to classify the sentiments. Table 4 shows the results of the sentiment model.

According to the TFIDF feature selection method, the results of the sentiment analysis classifiers showed that NB is the highest accuracy; 98% agreed with the results [38, 39]. In comparison, SVM 88% and the decision tree 71% agreed [40–42].

In [38], they proposed a Traveller Review Sentiment Classifier to analyze the traveler’s reviews on Egyptian Hotels and classify each sentiment based on hotel features. They used three classifiers, SVM, NB, and Decision tree, and it was found that NB was the highest accuracy. Results in [38] had shown that classifiers accuracy: NB, SVM, and Decision tree is 85%,79%, and 71%, respectively, which is very close to the results concluded in this research. The Citizen Sentiment model is a very good reflector of Citizen Intention since the Precision level was greater than 0.7 for the three classifiers.

Table 4. Citizen sentiment model results

Classifier	Accuracy	Precision	Recall
	TFIDF	TFIDF	TFIDF
SVM	88%	0.87	0.91
NB	98%	0.91	0.78
Decision tree	71%	0.84	0.85

Researchers in [39] presented a Job Candidate Rank model that analyzes the interviewers’ sentiment and calculates a rank for each job candidate. This approach analyses the interviewers’ sentiment, create a list of ranked keywords and then classifies the sentiments using machine-learning techniques. The results show that the Job Candidate Rank approach provides an accuracy of 93% when using the NB classifier and then SVM with a value of 89%, which also agrees with this research results.

In [40], they aim to define the factors that affect students’ decision to choose the university and predict student decisions based on testing cases using machine learning techniques. The results were convenient with the results shown in this research.

A classifier model for travelers’ intentions is designed in [41]. They classified the sentiments using Naïve Bayes and SVM, results had shown that the highest accuracy when using the Term Frequency Inverse Document Frequency feature selection method; however, the Support Vector Machine has the highest accuracy level when using the Bag of Words feature selection method. Results are aligned with the results of this research in which NB is the highest accuracy when using the TFIDF feature selection method.

After analyzing citizens’ sentiments, safety is the highest significant benefit of smart citizens that will affect their decision to be a resident, which is agreed with [2, 3]. In [2], Brazilian citizens in smart cities were motivated to go and live in the smart city to avoid crimes and kidnapping and to feel safe. Following the reduction of traffic congestion by 1082 sentiments which agreed with [4, 5], Citizens in Spain [4] face high traffic congestion on roads, leading them to go to the smart cities. Ireland’s smart cities followed the citizens’ comments by creating a survey that included their opinions of their reasons for living in smart cities. Most citizens talked about the traffic and deadlock problems, which agreed with the research results.

Positive energy sentiments were 1010, which agreed with [6] that tries to understand the factors affecting Japanese citizens to live in a smart city, green landscape 922, the clean place was 802. Technical facility is 700 sentiments. Table 5 categorizes the sentiment results into the ten previous categories discussed in Table 4.

Table 5. Citizen sentiment categories

Category	Count of sentiments
Reducing traffic congestion	1082
Increasing people safety	1210
Clean place	802
Reduce travel cost	292
Advanced technology	115
Parking availability	326
Technical facility	700
Green landscape	922
Reduce noise pollution	541
Positive energy and relaxation	1010
Total sentiments	7,000

5 Conclusion and Future work

Cities are crowded with people, and citizens face problems like traffic congestion, the risk to people’s safety, air pollution, travel cost, parking unavailability, technical Facilities problems, no Green landscape, Noise pollution, and negative Energy. With the evolution of technology, smart cities will be constructed with technical facilities that allow citizens to use these services easily. The Egyptian government is constructing smart cities with the theme of Egypt digital to allow its citizens to the new technology and solve the other problems raised by crowded places and high population. The research found that Egyptian citizen is motivated to move to smart cities once it is constructed. Governments should consider analyzing social media data to understand the citizens’ needs. Citizen engagement is the new trend that most governments are trying to follow. Citizen can express their opinions and ideas through social media, which will all governments to provide better services to citizens. Based on the above analysis and conclusions, the researchers believe that the digitalization of government services and the construction of smart cities will face a very high level of acceptance from Egyptian citizens while covering all urban problems stated in the research. The research recommends understanding citizen needs through social media, high level of citizen engagement, and a high response to citizen appeals. The research faces the following limitations: sentiments collected are limited to 7000 sentiments, three classifiers were tested, and only sentiments in English were processed. Future work can include Egyptian sentiments written in Arabic and franco arab; also gulf countries can be added to the sample size.

References

1. UN.: <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html>. Accessed 11 Nov 2022

2. Afonso, R.: Brazilian smart cities: using a maturity model to measure and compare inequality in cities, pp. 230–238 (2015)
3. Kwak, Y., Lee, J.: Toward sustainable smart city: lessons from 20 years of Korean programs. *IEEE Trans. Eng. Manage.* **70**, 740–754 (2021)
4. Appio, F., Lima, M., Paroutis, S.: Understanding smart cities: innovation ecosystems, technological advancements, and societal challenges. *Technol. Forecast. Soc. Change* **142**, 1–14 (2019)
5. Cardullo, P., Kitchin, R.: Being a ‘citizen’ in the smart city: up and down the scaffold of smart citizen participation in Dublin Ireland. *GeoJournal* **84**(1), 1–13 (2019)
6. Mizuno, Y., Odake, N.: Current status of smart systems and case studies of privacy protection platform for smart city in Japan, pp. 612–624 (2015)
7. Nam, T., Pardo, T.: Conceptualizing smart city with dimensions of technology, people, and institutions, In: *ACM International Conference Proceeding Series*, pp. 282–291 (2011)
8. Grady, M., Hare, G.: How Smart Is Your City?, pp. 3–5 (2012)
9. Alsabi, E., Alrwais, L., Alosimai, R., Alrwais, R., Dahanayake, A.: Benchmark for smart cities, *Front. Artif. Intell. Appl.* 280 (2016)
10. Yue, A., Mao, C., Chen, L., Liu, Z., Zhang, C., Li, Z.: Detecting changes in perceptions towards smart city on Chinese social media: a text mining and sentiment analysis. *Buildings* **12**, 1182 (2022)
11. Han, M., Kim, M.: A critical review of the smart city in relation to citizen adoption towards sustainable smart living. *HabitatInt.* **108**, 102312 (2021)
12. Adikari, A., Alahakoon, D.: Understanding citizens’ emotional pulse in a smart city using artificial intelligence. *IEEE Trans. Ind. Inform.* **17**, 2743–2751 (2021)
13. Kowalik, K.: Social media as a distribution of emotions, not participation. Polish exploratory study in the EU smart city communication context. *Cities* **108**, 102995 (2021)
14. Steinmetz, C., et al.: Liking, tweeting and posting: an analysis of community engagement through social media platforms. *Urban Policy Res.* **39**, 85–105 (2021)
15. Wang, A., Lin, W., Liu, B., Wang, H., Xu, H.: Does smart city construction improve the green utilization efficiency of urban land? *Land* **10**, 657 (2021)
16. Mendez Mena, D., Papapanagiotou, I., Yang, B.: Internet of things: survey on security. *Inf. Secur. J. Glob. Perspect.* **27**(3), 162–182 (2021)
17. Rasekh, A., Hassanzadeh, A., Mulchandani, S., Modi, S., Banks, M.: Smart water networks and cyber security (2021)
18. Schultz, W., Javey, S., Sorokina, A.: Smart water meters and data analytics decrease wasted water due to leaks. *J. Am. Water Works Ass.* **110**(11), E24–E30 (2021)
19. Phadtare, S., IndajeetJadhav, A.: Role of smart cities in sustainable development. *International J. Eng. Res. Technol.* 10(1). International Research Publication House (2017)
20. Habib, A., Alsmadi, D., Prybutok, V.: Factors that determine residents’ acceptance of smart city technologies. *Behav Inform Technol* **39**(6), 610–623 (2020)
21. Kuberkar, S., Singhal, T.: Factors influencing adoption intention of AI powered chatbot for public transport services within a smart city. *Int. J. Emerg. Technol. Learn.* **11**(3), 948–958 (2022)
22. Albreem, M.A., et al.: Green internet of things (IoT): an overview. In: *2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, pp. 1–6. IEEE (2017)
23. Al Ridhawi, I., Kotb, Y., Aloqaily, M., Jararweh, Y., Baker, T.: A profitable and energy-efficient cooperative fog solution for IoT services. *IEEE Trans. Industr. Inf.* **16**(5), 3578–3586 (2019)
24. Al-Zinati, M., Almasri, T., Alsmirat, M., Jararweh, Y.: Enabling multiple health security threats detection using mobile edge computing. *Simul. Model. Pract. Theory* **101**, 101957 (2020)

25. Shadroo, S., Rahmani, A.M., Rezaee, A.: The two-phase scheduling based on deep learning in the Internet of Things. *Comput. Netw.* **185**, 107684 (2021)
26. Ng, I., Wakenshaw, S.: The Internet-of-Things: review and research directions. *Int. J. Res. Mark.* **34**(1), 3–21 (2017)
27. Bibri, S., Krogstie, J.: Smart sustainable cities of the future: an extensive interdisciplinary literature review. *Sustain. Cities Soc.* **31**, 183–212 (2017)
28. Said, A., Tolba, A.: Accurate performance prediction of IoT communication systems for smart cities: an efficient. *Sustain. Cities Soc.* **69**, 102830 (2021)
29. Abou El Seoud, T.: How much Smart to be Smart?, *J. Urban Res.* 31 (2019)
30. Lopes, I., Oliveira, P.: Can a small city be considered a smart city. *Procedia Comput. Sci.* **121**, 617–624 (2017)
31. Batty, M.: Big data, smart cities and city planning. *J. Dialogues Hum. Geogr.* **3**, 274–279 (2013)
32. (2022). Gopp <http://gopp.gov.eg/>
33. NUCA: Accomplishments: Building new fourth-generation cities. New Urban Communities Authority, Ministry of Housing, Utilities & Urban Communities (2018). <http://www.lands.nuca.gov.eg>
34. Abou El Seoud, T.Z.A.: Towards sustainability: smart cities in the Egyptian environment. *J. Urban Res.* **31**, 123–142 2019
35. Zanella, A., Bui, N., Castellani, A., Vangelista, L., Zorzi, M.: Internet of things for smart cities. *IEEE Internet Things J.* **1**(1), 22–32 (2022)
36. NutifafaArku, R., Adrian Buttazzoni, A., Agyapon-Ntra, K., ElmondBandauko, A.: Highlighting smart city mirages in public perceptions: a Twitter sentiment analysis of four African smart city projects. *Cities* **130**, 103857 (2022)
37. Pawełozek, I.: Towards a smart city—the study of car-sharing services in Poland. *Energies* **15**, 8459 (2022)
38. Mostafa, L.: Machine Learning-Based Sentiment Analysis for Analyzing the Travelers Reviews on Egyptian Hotels, March 2020, In book: Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020). https://doi.org/10.1007/978-3-030-44289-7_38
39. Mostafa, L., Beshir, S.: Job Candidate Rank Approach using Machine Learning Techniques (AMLT) (2021)
40. Mostafa L., Beshir S.: University Selection Model Using Machine Learning Techniques. In: Hassanien A.E. et al. (eds.) Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2021)
41. Mostafa, L., Beshir, S.: Understating factors affecting traveling during covid-19 using sentiment analysis. In: Hassanien, A.E., Snášel, V., Chang, K.-C., Darwish, A., Gaber, T. (eds.) AISI 2021. LNDECT, vol. 100, pp. 105–114. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-89701-7_10
42. Mostafa, L., Beshir, S.: Using Gamification in Egyptian E-Government. In: Hassanien, A.E., Snášel, V., Tang, M., Sung, T.W., Chang, K.C. (eds.) AISI 2022. LNDEC Technologies, vol 152, 344–353. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-20601-6_31
43. Knime (2022) <http://www.knime.com/>. Accessed 11 Nov 2022



Toward a Generative Chatbot for an OER Recommender System Designed for the Teaching Community: General Architecture and Technical Components

Sandoussi Rima¹(✉), Hnida Meriem^{2,3}, Daoudi Najima^{1,2}, and Ajhoun Rachida¹

¹ ENSIAS, Mohammed V University in Rabat, Rabat, Morocco

² School of Information Sciences (ESI), Rabat, Morocco

³ Mohammadia School of Engineers (EMI), Mohammed V University in Rabat, Rabat, Morocco

Abstract. A chatbot (Chatting Robot) is a computer program that enables people to communicate with machines using everyday language. The use of chatbots evolved quickly in the last years for different domains, including education and teaching. This paper presents related research conducted in the past year to contextualize the research. Next, we present the general context of a chatbot, its different intrinsic components, and their general architecture to constitute a solid base for the proposed proper system.

Furthermore, we present the architecture of the proposed Chatbot by describing, in-depth, the function of every component and detailing the system's scenario. This study is a solid foundation for the final generative Chatbot for Open Educational Resources (OER) Recommender System. This paper aims to present the first step of developing a OER Chatbot" Recommender designated to the teaching community, which is the general architecture of the global system's first component.

Keywords: Chatbot · Generative chatbot · Open Educational Resources · General architecture · Technical components

1 Introduction

E-learning has witnessed an important breakthrough in the era of the artificial intelligence revolution. Several learning platforms were designed to offer high-quality and personalized services to learners. However, two main challenges are still encountered. On the one hand, the production of educational resources is expensive and difficult; on the other hand, the services to authors are still considered elementary. Open educational resources (OER) can be considered an efficient solution to address educational content. Regarding the author's services, OER reusing requires developing interactive and personalized services for content authors.

Chatbots are one of the most promising artificial intelligence technologies enhancing interactivity between human users and machines. Chatbots are "computer programs that

interact with users using natural languages [1]”. They are intelligent agents with whom users can communicate, typically via text or speech, to be given an automatic reply to a specific query. Developing a generative chatbot is a suitable solution to assist course authors during the production process of new educational content to find the most relevant content. The literature mentioned two categories of chatbots: Rule-based and generative chatbots.

The first category uses knowledge-based containing predefined questions and answers. The Chatbot’s knowledge is hand-coded by humans and organized and presented using conversational patterns. A larger rule database allows the Chatbot to respond to a wider range of user input. However, this model is not resistant to grammatical and spelling errors in user input. The second one generates language combinations rather than selecting from a list of predefined responses. To model and train the Chatbot system, these models employ mainly Natural Language Understanding (NLU) and natural language generation (NLG) as part of natural language processing (NLP) and machine learning and deep Learning techniques.

This paper aims to outline the overall chatbot architecture to personalize Educators’ services to reuse OERs. The proposed Chatbot is generative for an educator-focused recommender system that generates customized lists of recommended OERs based on the tutor’s requests. The first section is dedicated to related works, and the second section focuses on the generative model by presenting an overview of the general framework of a generative chatbot with its various components. The third section is dedicated to our own system’s architecture.

2 Context and Problem Statement

2.1 Conversational Agents and Chatbots

Conversational agent systems have become more popular in various industries, including business, tourism, and education. Chatbot development models have evolved from traditional rule-based to retrieval-based models, then to generative models, to increase accuracy and make conversations more realistic and reliable [2]. In 1950, turning questioned if a computer program could converse with people without them realizing it was a computer program; this question is known as the Turing test and was considered to be the procreative idea of chatbots. The first Chatbot ever made named Eliza, it was created in 1966, and it “is a program which makes natural language conversation with a computer possible. [3].

The chatbot architecture remains the same regardless of how simple or complex the Chatbot is. The front end is where the user interacts with the bot, and the NLP Engine processes the responses and generates the appropriate response. All the types of the above chatbots should include the following components:

- (A) **User interface:** Conversational user interfaces (CUIs) are a chatbot’s front end that embodies the conversation’s physical representation. They are divided into two types: text-based and voice-based assistants. [4]

- (B) **Natural language processing (NLP)** [5]: allows chatbots to convert users' messages (Text or voice) into structured data that a machine can understand. NLP could be made up of the following steps:

(B1) *Tokenization*, also known as linguistic analysis, divides a sentence's string of words into smaller "tokens" based on their meaning and relationship to the whole sentence.

(B2) *Normalization*, also known as syntactic analysis, is the process of identifying and correcting words.

(B3) *Entity recognition* is the process of searching for keywords in order to determine the topic of a conversation.

(B4) *Semantic analysis* is the process of inferring a sentence's meaning by understanding the meaning of each word and its relationship to the overall structure.

- (C) **Natural language understanding (NLU)**: is a subfield of NLP that focuses on recognizing patterns in unstructured speech input to understand the meaning of human speech, and it is composed of a Dictionary to determine the meaning of a word, a Parser to determine if the syntax of the text conforms to the rules of the language, and Grammar rules to break down the input based on sentence structure and punctuation. NLU enables chatbots to classify users' intents and generate a response based on training data.
- (D) **Knowledge Base**: is a library of information that the Chatbot relies on to retrieve data for responding to users. Knowledge bases differ depending on the needs. Some chatbots are integrated with web scrapers to collect and display data from online resources.
- (E) **Data storage** (SQL or cloud) to store and save conversations for user's uses and bot training and testing purposes.
- (F) **Dialog manager** is the component in charge of the flow of the conversation between the user and the Chatbot; it keeps track of the interactions within the conversation to determine how to respond.
- (G) **Natural language generation (NLG)** to transform machine-produced structured data into human-readable data.

2.2 Problem Statement

The research about OER recommenders is oriented toward making appropriate predicted recommendations that are context-based and take into account the users' preferences and tendencies. However, learners and students are primarily the focus of the existing systems, resulting in a scarcity of OERs recommenders dedicated to teaching purposes. Based on the system of the research [6], and to make a solid continuity to it, we aim to add a generative chatbot as a novel and efficient recommendation technology; that constitutes an excellent tool for addressing the choosing process issue for the appropriate resources, which is a common task in the teaching practice.

Objectives. To develop the general architecture of the Chatbot as a digital tool with personalized Recommendation lists of OERs to support educators in course creation and to present the recommendation scenario of the Chatbot.

3 Theoretical Foundation and Related Works

Several researchers have been concerned about using generative Chatbots in e-learning in recent years. In [7], the authors assume that chatbots in education have received little attention. They then define their Chatbot using Xatkit, a chatbot development framework, to finally propose an Encoder-Decoder framework for intent recognition to detect the user's intent. This Chatbot will be tested in the field of education to improve and simplify professors' teaching and students' learning, as well as to reduce faculty burnout and increase comprehension speed. In [8], the authors propose a chatbot that is effective for students to learn. It can respond to questions at any time without requiring human interaction. It is created using natural language processing and deep learning technologies. As a result, the Chatbot can understand questions at a higher level. Simple to use, appropriate for people of all ages, and provides the most thorough answers to questions. This Chatbot deploys a long short-term memory algorithm. In [9], the authors focus on the language-related challenges and roadblocks that prevent Arabic chatbots from being implemented. This paper also includes basic guidelines for implementing effective Arabic chatbots for academic purposes. In [10], the authors developed a chatbot used on Facebook for the official admission Fan page of the National Vietnamese Economics University. A chatbot that provides students with daily curriculum updates, new student admissions, tuition fees, and IELTS writing task II scores, among other things. With an accuracy of 97.1 percent on a test set, this model can detect more than fifty questions from users' input. In [11], the authors developed a chatbot that improves students' research knowledge, resulting in positive learning outcomes and a more personalized learning experience. The research goals are to create the Chatbot, assess its effectiveness, and investigate its effects on students' knowledge. Experts give the Chatbot a very high rating for applicability ($= 4.67$, $SD = 0.08$), with suggestions to add more research content and interactive learning. In [12], the authors implement a text chatbot feature, named *lixa*, to Homelab, a discussion platform on course material and assignments for students. The aim is to assist users in understanding the app and guiding them through its use. To help it converse with users smoothly, the Chatbot uses a multilayer perceptron combined with a special text dataset. In [13], a knowledge-based chatbot system was integrated into the teaching activities of a physical examination course. The findings suggest that chatbots have a lot of potential for use in nursing training courses. In [14], authors developed a new method for extracting keywords using Natural Language Processing (NLP) techniques. This paper aims to create an e-learning AI-powered Chatbot that allows learners to interact with it. The Chatbot's core is built on indexed multimedia content, allowing it to search for information quickly.

4 Global System Proposition for Course Authors During the Educational Content Production Process

This section presents the Recommender system that customizes and generates a list of OERs to meet the educators' most specific preferences. This system provides a genuine recommendation scenario based on a chatbot with a recommendation process for an educator. It operates as an interactive modular component that could be embedded in other

e-learning platforms. The OER recommendation is formulated through an interactive exchange between the educator and the Chatbot. The Chatbot asks questions to understand the teacher's most specific and individual needs, then refines the recommendation to give a complete list of appropriate OERs. The rating of the proposed recommendation is the process that will be used to enhance the recommendations' quality and precision and to update the ranked list of OERs and the educator preferences for further and future recommendations.

Once the educator joins the system by signing up to get personalized OERs recommendations, building the profile begins. The profile data storage is updated depending on the models and characteristics established in our solution. It is mainly based on intrinsic data to the registration provided by the teacher on the sign-in form, such as the category and level of the students, the subject to be taught, and the type of course. Then the profile is continuously updated with data related to the preferences and behaviors collected as the user browses and makes actions in the system.

The recommendation is based on an assisted automated process founded on a chatbot-controlled interaction process. The proposed Chatbot is a generative conversational artificial intelligence that asks questions using a learning algorithm to guide and orient it. This conversation aims to determine the users' most precise demands. In general, the Chatbot should deliver more exact and tailored results.

4.1 Chatbot Requirements and General Architecture

Messenger Interface: Enables queries to be sent and received by the users.

Natural Language Understanding NLU; Users communicate with Chatbot using natural language. The Chatbot we develop must be able to comprehend user messages and gather the required data. The Chatbot's NLU capacities can be attained through intent classification and entity extraction. And for its configuration, this component needs the intent list, entity list, and list of potential entities and synonyms, also the training data and labels to be defined.

Intent Classification: A collection of utterances with a similar meaning is called intent. The intent classification collects groups and classifies those utterances. The intent classifier analyzes user input, determines its significance, and associates it with one of the intents supported by the Chatbot.

Entity Extraction: Contrary to intents, entities are utterances with unique meanings, such as colors, shapes, and names. So, Entities typically have a proper name, and they differ from entities as a whole in that they always signify the same notion. Entity extraction discovers, prepares, and categorizes entities found in raw text data and is responsible for extracting key information from the users' queries.

The actions environment is the part of the Chatbot that links all the components and is responsible for the allocation and execution of the actions and their programs.

The main core of the action environment is the action definition and allocation, which is responsible for outlining and coordinating the asynchrony of the bot actions. In other words, this component constitutes a traffic controller to maintain the chatbot actions order. We define a program function in this component for each action defined in the configuration data. If the action is performed, the function will be invoked. The function's body can contain an algorithm, steps, or calls to another component's function [15].

The link between the Chatbot and other external systems is performed via APIs.

The Dialog Manager. Copes with any change in the conversation. The dialog manager interferes whenever the user interrupts the query with a new one [16].

NLG: With the NLU, this component constitutes the NLP component, and it is responsible for the automatic generation of the response to be comprehensive by the user.

QA System [17]: The Chatbot generates an utterance as a response for the user. This component is composed of the following items:

- (A) **Response generator:** This component generates personalized utterances for the users. The utterances are built after the data training is effectuated
- (B) **Automated training:** The component responsible for training data to generate an utterance
- (C) **Knowledge base:** This component is an information library on which the Chatbot relies to retrieve data to respond to users.

Figure 1 illustrates the system's overall architecture, including the Chatbot.

4.2 Global System Scenarios

As we see in Fig. 2, the educator enters a query via the messenger interface. The messenger is connected to the NLU component. The NLU is composed of the intent classification, which focuses on predicting the intent of the query and the extract entities that extract particular information from unstructured text and organize it into predetermined categories.

The NLU component intercepts the input query from the user that transforms it into structured data understandable by the system. The structured data is distributed to the action definitions and allocations in the action environment; this component is responsible for allocating and diffusing the data and deciding the next right component. The structured data go to the QA system and search for suitable answers in the knowledge base.

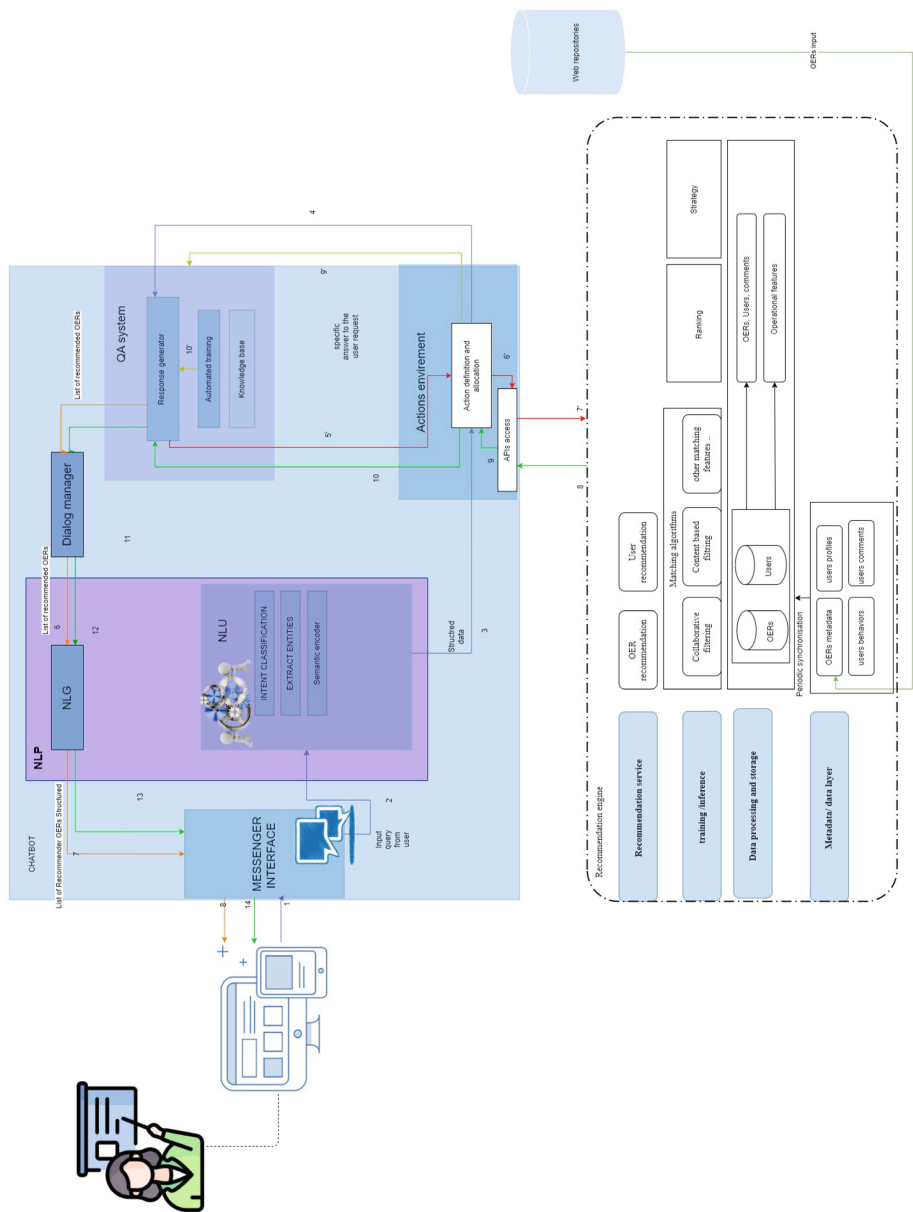


Fig. 1. The general architecture of the system and its components

If the answer exists already in the knowledge base, it will be communicated to the NLG, then to the messenger interface, and getting feedback from the user in the end.

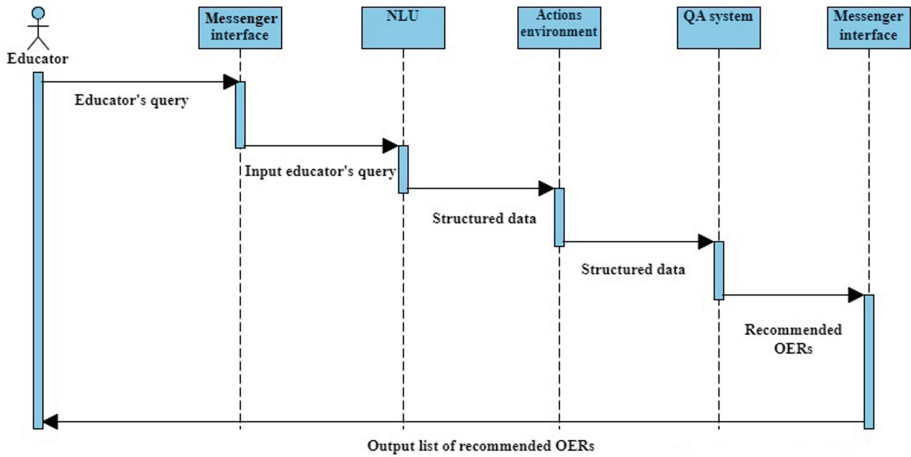


Fig. 2. The first scenario of the system

Suppose the answer does not exist in the knowledge base. In that case, the data will be returned to the definition of the action, then to the API, which is the component that integrates every system's API, including the one that assures the communication between the Chatbot and the recommender engine. The recommender engine gives recommendations that suit the educator's query. This recommendation is sent to the API, then returns to the definition of the action and updates the knowledge base for future uses. Once the knowledge base is updated, the OERs are communicated to the NLG and given to the educator (refer to Fig. 3).

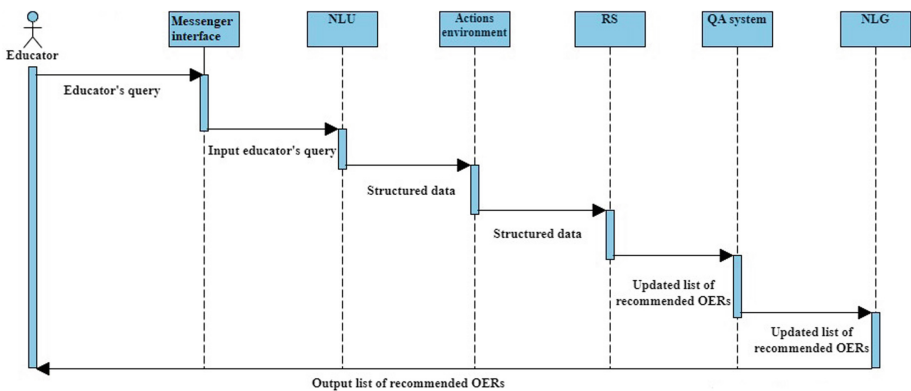


Fig. 3. The second scenario of the system

5 Conclusion

This paper is an overview of a general architecture of a chatbot for an OER Recommender System designated to the teaching community and the course creators. To commence constructing a chatbot requires knowledge of a wide range of AI issues and the integration availability of components. This research exhibits our architecture with exhaustive explication and detailed components, and it constitutes the first result of our previous systematic literature review [18]. Unlike several chatbots and systems in the education field that are students' or learners' oriented, this architecture is for a teaching-oriented Chatbot that will be embedded in a recommender system for OER for teachers and course makers. Further work of this research would be exploring in detail the components of our architecture to build up our Chatbot by implementing ideas presented in this research, which is thus a significant step toward building our final system.

References

1. Shawar, B.A., Atwell, E.: Chatbots: are they really useful? *LDV Forum* **22**, 29–49 (2007)
2. Caldarini, G., Jaf, S., McGarry, K.: A literature survey of recent advances in chatbots. *Information (Basel)* **13**(1), 41 (2022)
3. Lin, Z., et al.: CAiRE: An end-to-end empathetic chatbot. *Proc. Conf. AAAI Artif. Intell.* **34**(09), 13622–13623 (2020)
4. Khurana, D., Koli, A., Khatter, K., Singh, S.: Natural language processing: state of the art, current trends, and challenges. *arXiv [cs.CL]* (2017)
5. Theosaksomo, D., Widyanoro, D.H.: Conversational recommender system chatbot based on functional requirement. In: 2019 IEEE 13th International Conference on Telecommunication Systems, Services, and Applications (TSSA) (2019)
6. Faqih, B., Daoudi, N., Hilal, I., Ajhoun, R.: Pedagogical resources indexation based on ontology in intelligent recommendation system for contents production in d-learning environment. *J. Comput. Sci.* **16**(7), 936–949 (2020)
7. Baha, T.A., Hajji, M.E.L., Es-Saady, Y., Fadili, H.: Towards highly adaptive Edu-Chatbot. *Procedia Comput. Sci.* **198**, 397–403 (2022)
8. Kasthuri, E., Balaji, S.: Natural language processing and deep learning chatbot using long short term memory algorithm. *Mater. Today: Proc.* (2021). <https://doi.org/10.1016/j.matpr.2021.04.154>
9. Almurayh, A.: The challenges of using arabic chatbot in Saudi universities. *IAENG Int. J. Comput. Sci.* **48** (2021)
10. Nguyen, T.T., Le, A.D., Hoang, H.T., Nguyen, T.: NEU-chatbot: chatbot for admission of national economics university. *Comput. Educ.: Artif. Intell.* **2**, 100036 (2021)
11. Vanichvasin, P.: Chatbot development as a digital learning tool to increase students' research knowledge. *Int. Educ. Stud.* **14**(2), 44 (2021)
12. Aqil, A.N., Dirgantara, B., Ahmad, U.A.I., Septiawan, R.R., Suherman, A.L.: Robot chat system (chatbot) to help users' homelab' based in deep learning. *Int. J. Adv. Comput. Sci. Appl.* **12**(8), 1–6 (2021)
13. Chang, C.-Y., Kuo, S., Hwang, G.: Chatbot-facilitated nursing education: incorporating a knowledge-based chatbot system into a nursing training program. *Educ. Technol. Soc.* **25**, 15–27 (2022)
14. Janati, S.E., Maach, A., El, D.: Adaptive e-learning AI-powered chatbot based on multimedia indexing. *Int. J. Adv. Comput. Sci. Appl.* **11**(12), 1–10 (2020)

15. Mauldin, M.L.: Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition. *AAAI* **94**, 16–21 (1994)
16. McTear, M.: The role of spoken dialogue in user–environment interaction. In: *Human-Centric Interfaces for Ambient Intelligence*, pp. 225–254. Elsevier (2010)
17. Chandra, Y.W., Suyanto, S.: Indonesian chatbot of university admission using a question answering system based on sequence-to-sequence model. *Procedia Comput. Sci.* **157**, 367–374 (2019)
18. Sandoussi, R., Hnida, M., Daoudi, N., Ajhoun, R.: Systematic literature review on open educational resources recommender systems. *Int. J. Interact. Mobile Technol.* **16**(18), 44–77 (2022)



An Emoticon-Based Sentiment Aggregation on Metaverse Related Tweets

Mousumi Bhattacharyya¹(✉), Asmita Roy¹, Sadip Midya¹, Anirban Mitra¹,
Anupam Ghosh², and Sudipta Roy³

¹ Department of Computer Science and Engineering, Sister Nivedita University, DG Block
(New Town), Action Area I, 1/2, New Town, West Bengal 700156, India
{mousumi.b, asmita.r, sadip.m, anirban.m}@snuniv.ac.in

² Department of Computer Science and Engineering, Netaji Subhash Engineering College,
Techno City, Panchpota, Garia, Kolkata, West Bengal 700152, India

³ Artificial Intelligence and Data Science, Jio Institute, Navi Mumbai 410206, India
sudipta1.roy@jionstitute.edu.in

Abstract. The metaverse is referred to as a virtual reality space, integrating the real world with the digital world. People have gained interest in this as many well-known companies like Facebook are entering this world of the metaverse. Most people are curious about the possibilities and future of the metaverse considering the various prospective applications it is going to offer like virtual meetings, learning environments and many more. However, there are also worries about potential negative effects as in the metaverse signals from the real world can be transmitted in the virtual world. In this regard, people express their feelings on various social media platforms. The various data obtained from this social media platform is a popular study subject as it gives an important insight into the thoughts of society towards a new event. Thus, in this work twitter, which is one of the extensive social media platforms internationally used to express thoughts, is chosen as the platform of study. Tweets are analyzed to assess people's attitudes and tendencies towards an event. In the present generation, people express themselves not only with words but also emojis. Therefore, in this paper tweets comprising of both text and emojis are considered for analyzing human sentiments towards the metaverse and the potential risk associated with it. The dataset is comprises of tweets having the word the metaverse and hashtags the metaverse. Observed analysis shows an improvement of over 2.92% for the proposed method.

Keywords: The metaverse · Sentiment analysis · Machine learning · Tweet analysis

1 Introduction

With the advancement of technologies and the growth of the internet, the concept of the metaverse which was coined in the 1990s is becoming applicable in real life at a large scale. The metaverse is referred to as an immersive universal world of the Internet which is facilitated by virtual and augmented reality. The advancement of ubiquitous computing

with its accessible hardware infrastructure has gained popularity for the concept of the metaverse. The realm of the metaverse allows humans to exist mentally in the virtual universe. This will allow humans to experience the feeling of reality in the digital universe while physically being in any environment. The metaverse not only has the capacity of replicating the physical world virtually but can create various digital layers in this replication. The advancement of this virtual world is reaching a stage where the metaverse can be created only by processing human thoughts. It is also predicted with the growth of the metaverse that 25% of humans will spend at least an hour in the virtual universe by 2026 [1]. This point of the virtual universe brings up two different perspectives – one where this universe is the next step for the Internet having great developmental opportunities for humanity. The amalgamation of the physical and virtual worlds brings the likelihood of humans realizing their dreams without affecting others. Along with this the being of immortality in this digital world also attracts people. The other one brings a threat to the physical world with a possibility that might end humanity. Severe reactions are observed that bring in a threat to the physical world. Therefore, the advent of the metaverse brings with it a positive as well as negative impact on society. Thus, it is very much necessary to predict the future of the metaverse and its potential vulnerabilities [2]. The digital world is an integral part of human life in today's world. Preventing the negative impact of the metaverse on society requires an imperative analysis of the interactions between the physical and the virtual world. In the current technological era, social networks are immensely used by humans. These social networking sites help people express their thoughts, ideas, events, news, and situation about their daily life. Thus, examining the thoughts expressed by people on social media networks will help in determining the adversities associated with the metaverse.

2 Motivation and Contribution

Most of the sentiment analysis works on text and ignores the emoticons; which leads to partially inappropriate results. As data is supplied in the form of reviews, photos, emoticons, and videos in the present generation. Emoticons are symbolic expressions made up of symbols like “:”, “=”, “-”, “()”, or “(” and frequently depict facial expressions. Emoticons can be read either way, such as “:-(” for a sad face or “(” for a cheerful one. Monitoring these emoticon symbols in conjunction with the text is crucial for identifying the true feelings, such as happiness, frustration, rage, grief, etc., which are then divided into positive, negative, and neutral categories. Also, with the advancement of technology, the amalgamation of real space to virtual space leading to a world of the metaverse, urges a need to analyze human sentiment about the new event. This analysis will result in partial findings if only text-based tweets are considered. Thus emoticon-based tweet analysis is the need of the hour to gain better accuracy. This motivated the authors to develop an emoticon-based sentiment aggregation algorithm to predict positive negative, or neutral feelings. The contribution of the proposed approach are as follows –

- a) Emojis are extracted from tweets and mapped to emoticon data set and corresponding score values for emoticons are set for efficient analysis.
- b) The proposed model shows increased accuracy, which will be better applicable in online conversation.

- c) Accuracy being a pertinent measure of e-ciency the outcome is varied using AUC-ROC model.

3 Related Work

The metaverse, according to Stephenson, is a virtual reality setting where individuals can communicate with one another in a setting created by a computer. The signals from people's bodies in the actual world are picked up by physical equipment on the Internet to be sent to the virtual world. This is how the real world and the metaverse are connected [3]. In [4, 5], authors proposed methods to interact between physical and virtual worlds. For customizing virtual characters in [6] authors developed a user interface.

The above techniques mainly focused on how the virtual and real-world interaction will be done. Furthermore, analyzing human emotions associated with the metaverse is required [7]. The advent of technology and artificial intelligence has increased emotional online communication, where people communicate about personal and emotional issues on social media platforms [8].

The general SA researcher deals with either text or emoticons. Most social media research in SA has been conducted on text using machine learning (ML) techniques. However, due to a lack of resources and the difficulty of emoticons, SA on both text and emoticons has been mostly disregarded [9]. The most popular emoticons are ranked higher, the least popular ones are ranked lower, and others are ranked in the middle. On the other hand, emojis can be used in groups and at the end of tweets. The algorithm will use the discrete probability distribution formula to find the number of emoticon occurrences in the tweets as well as the emotion label c in the sentiment distribution for the set of pertinent tweets [10].

User effectively tags their text with an emotional state when they employ an emoticon (emoji) in electronic communication. The emoticons-based sentiment analysis is a great technique due to a number of aspects. The emoticons and emojis can first be used with data that is multilingual. It is quite simple to use emojis and emoticons on text in any language when you have access to a dictionary. Additionally, to improve outcomes, the emoticons-based analysis can be coupled with the current lexicon and machine learning strategy [11].

Emojis are frequently used by web users or chat social networks to express the emotions that underlie their textual interactions to make their messages easier to understand, especially for brief words. Any chat or social network that uses semantic-based context recognition techniques may understand text-based emoticons and convert them into audio data. On the other hand, image-based emojis require image-recognition algorithms in order to fully comprehend the semantic context [12].

When creating features for comparison with those that were directly created from emoji-label words, emojis were additionally translated into equivalent sentiment words. Additionally, taking into account the various purposes of emojis in texts, all posts in the dataset are classified according to their emoji usage and looked at how the algorithm's performance changed over time. Emojis work well as expanding features to raise the precision of sentiment analysis algorithms, and by taking varied emoji usages into account, the algorithm performance can be further improved [13].

4 Proposed Emoticon Sentiment Aggregation Approach

Any prediction model's efficiency depends on how the data set is prepared. In this paper, not only tweets comprising the term the metaverse are analyzed but also the related emojis with the corresponding tweets are considered. In case of the absence of a metaverse term, the associated tweet's classification is considered with precedence. In this context analysis of actual tweets is very important and precious as it guides towards the sentiment of Twitter. A labeled dataset is the primary prerequisite for building a reliable Machine Learning (ML) model. The ML model can forecast sentiment for given unseen text by learning different patterns from the dataset. The classification of human sentiment based on tweets made on the metaverse is shown in Fig. 1. The actions listed below must be taken in order to train a unique sentiment analysis model.

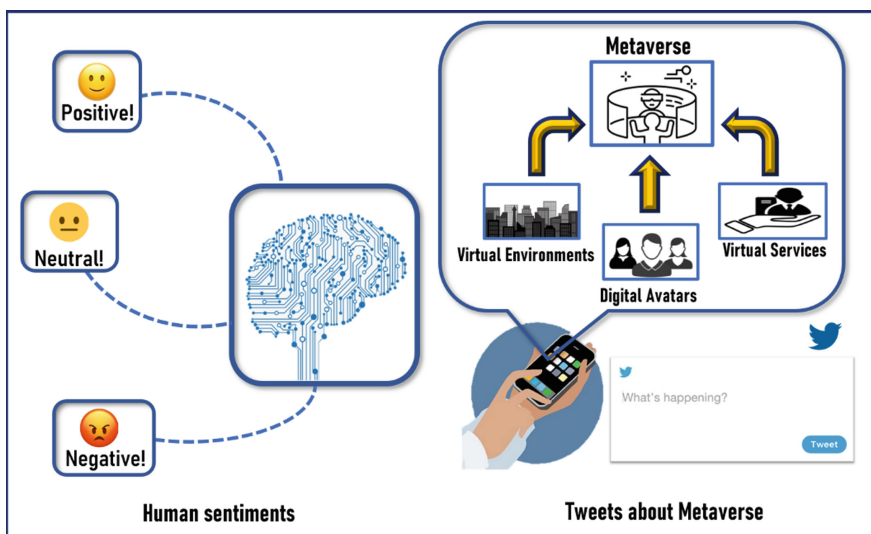


Fig. 1. Classification of human sentiment based on tweets made on the metaverse

Collection of Data: A raw, labelled dataset has been downloaded from Kaggle. Using the social media API offered by many social networking sites, An Emoticon based Sentiment Aggregation on the metaverse related Tweets 5 the social media data, specific data from Twitter, are harvested on a big scale. The description of the data field is given in Table 1. When extracting data, metaverse-related hashtags and keywords are found. Based on the length of time and volume of data extracted, the API establishes limits on the data access rate. A generic data interchange format, such as CSV, will be used to extract and store all pertinent data based on the query in the centralized storage location (Comma Separated Value). A broad range of cleaning factors must be applied to the data obtained from various social media platforms to obtain a well-defined set of data. These tweets were gathered using a Python script and the Twitter API. To gather more tweet samples, a daily query for the high-frequency hashtag #The metaverse is executed for a predetermined amount of time.

Table 1. Dataset

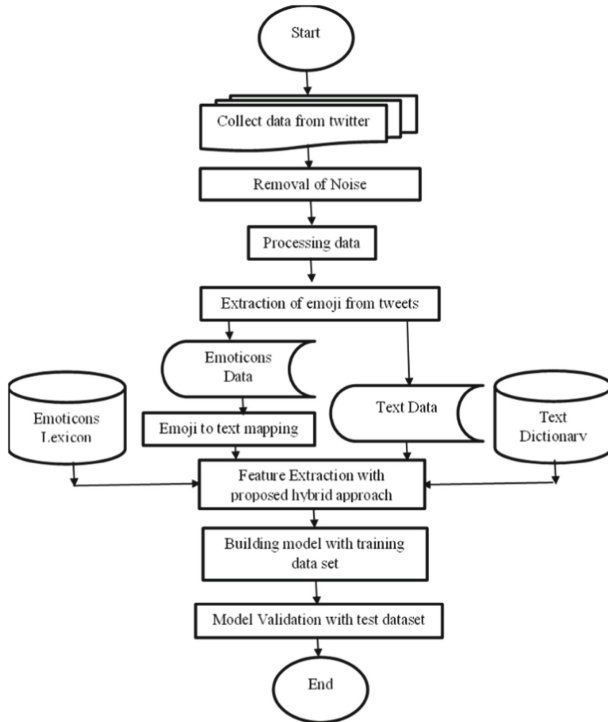
Column Name	Description
User_ame	Name of the account holder
User_location	Location of the account holder
User_created	type of the person; i.e.: student, professional etc
Use_followers	Creation date of the account
User_friends	Followers of the account holder
User_favourites	Friends of the account holder
User_verified	Favorites of the account holder
date	Verified account holder or not
text	Date of the tweet
hashtags	Text of the tweet
source	Hashtags #The metaverse
is_retweet	Source device of the Tweet

Preprocessing Text: Tweets consist of parts that are not relevant for analysis. Removal of such parts and tokenization of the tweets are required to get better results. Noise Removal - There is a lot of noise and useless terms in the social media data that has been extracted. Pre-processing and data cleansing must be done properly. Hyperlinks, punctuation, digits, stop words, and emojis are among the things that are removed during the noise-removal process.

1. **Tokenization:** This reduces the entire tweet to words, by eliminating all punctuation, including commas and blank spaces. By separating the incoming data into pertinent words, tokens are produced. These tokens are then used as input for further processing.
2. **Lemmatization:** In this process fully eliminating or switching out a word's su-x is carried out. This process produces a lemma. In the case of the word "go," lemmatization continues as "going, went, gone, goes," and "go" alone is picked for pluralization
3. **Corpus Formation:** A tweet-wise vector composed of lemmas is produced after all the tweets are lemmatized. While corpus refers to the entire group of vectors, the document refers to each vector separately. The corpus's features are extracted in the following step. In this context, a document is any piece of natural language text that is at least one coherent paragraph in length and can be as short as a one-line tweet or as long as a news article, blog post, or forum message.

Proposed Sentiment Analysis: In the proposed work sentiment analysis is done by considering tweet data as well as emojis with the tweets. This requires extracting the emojis from the tweet and then further scoring the text data and emoji into the same level. The following process explains the proposed methodology.

Input:	A document set $D = \{D_i\}$ where $i=1$ to A	//set of all documents or tweets
	A text set $T_i = \{t_k\}$ where $i = 1$ to A , $k = 1$ to $a1$	//set of metaverse related texts in document D_i
	A emoji set $E_i = \{e_l\}$ where $i = 1$ to A , $l = 1$ to $a2$	//set of metaverse related emojis in document D_i
Output:	Score Set $S = \{S_i\}$	//Set of scores of each document D_i
Algorithm		
Step 1:	for $\forall D_i$ in D do	
Step 2:	If $T_i <> \{\emptyset\}$ && $E_i <> \{\emptyset\}$ then	
Step 3:	if $S(T_i) == S(E_i)$ then	
Step 4:	$S_i = S(T_i)$;	
Step 5:	else	
Step 6:	$S_i = 0.7 * S(T_i) + 0.3 * S(E_i)$;	
Step 7:	else if $T_i == \{\emptyset\}$ then	
Step 8:	$W_{ij} = \{\text{words extracted from } E_i\}$, where $ W_{ij} = m$;	//m is the no. of words extracted from E_i
Step 9:	$S_i = (\sum_{j=1}^m S(W_j)) / m$;	
Step 10:	else if $T_i <> \{\emptyset\}$ then $E_i == \{\emptyset\}$ then	
Step 11:	$S_i = S(T_i)$;	
Step 12:	done	

Fig. 2. Algorithm of the proposed model**Fig. 3.** Flowchart of the proposed model

1. **Separation of emojis from tweets:** We load a document first before performing sentiment analysis on it. The material is then separated into paragraphs or sentences, depending on the document. We check for emoticons as well as punctuation signs like “.”, “!”, “,” and “?” when dividing a document into sentences because most emoticons are placed at the conclusion of a text segment. After separating text and emoji, emoji are converted to text. Then at the document level, the results are merged using sentiment aggregation. The presence of emoticons is checked in every text segment. A list of character sequences that represent emoticons and their corresponding sentiment ratings, known as an emoticon sentiment lexicon, can then be employed. Words are regarded in this context as character sequences that are divided up by whitespace characters. A text segment’s sentiment is assessed based on the emotion imposed on the text by its emoticons if a word in the segment matches a character sequence in the emoticon sentiment lexicon; otherwise, the segment is examined for the emotion expressed by its sentiment-conveying words.
2. **Feature Extraction using sentiment aggregation:** All text segments are recombined into a single document after each text segment’s sentiment has been identified. A document may contain both sections with and without emoticons, it should be noted. The document sentiment score is then determined as a weighted average of all segment-level sentiment scores, where the weights reflect the proportions of the number of emoticons or words (as appropriate) in each segment that are associated with the corresponding sentiment. A traditional bag of word models can be used for text feature extraction. As this model is adaptable and applicable in a variety of ways to extract features from documents. The assumption is that if two documents contain comparable information, then they are similar. Furthermore, something about the document’s significance can be inferred from its content. The goal is to convert each free-text document into a vector that can be used as either the input or output for a machine-learning model. The proposed algorithm is shown in Fig. 2 and the flowchart is given in Fig. 3.

5 Result Analysis

Naive Bayes and Support Vector Machine (SVM) is used for training the prediction model in the proposed approach. SVM is chosen as it is memory efficient and effective in high dimensional spaces. And Naive Bayes approach is highly scalable in terms of the number of predictors which make fast and real-time predictions. The proposed prediction model is expected to process several tweets at any point in time considering the extensive use of social media in today’s world. Thus, justifying the selection of the SVM and Naive Bayes approach. 70% of the total relevant corpus is used for training the model; whereas the remaining 30% is used for testing the model.

A series of experiments are run to validate the effectiveness of emoticon-based sentiment analysis. In order to establish a baseline for comparison, we first evaluate how well the proposed framework performs without considering the information that is conveyed by emoticons. This effectively reduces the functionality of our pipeline to that of a cutting-edge lexicon-based approach to document-level sentiment analysis. The effectiveness of the proposed sentiment analysis methodologies is examined on the test

collection, in which each document has at least one emoticon, in order to gain a clear understanding of the influence of sentiments given by emoticons. In Table 2 the number of total tweets, relevant tweets and rejected tweets is shown (Fig. 4).

Table 2. Analysis of collected tweets

Total tweets	50001
Tweets having the metaverse	45317
Rejected Tweets	4684

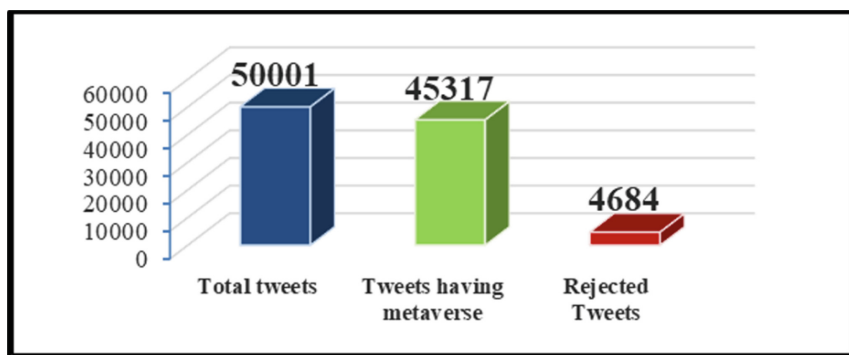


Fig. 4. Total tweets, Relevant tweets and Rejected tweets

Table 3 and Fig. 5 show the comparison between text sentiment analysis and our proposed sentiment aggregation approach with text and emoji both.

Using AUC-ROC the outcome is verified. For classification, there are various unique evaluation measures. Accuracy is the most popular metric because it measures how frequently the model is correct. This metric is great since it is straightforward to comprehend and often calls for the most precise estimates. The AUC of our model is displayed in Fig. 6.

Table 3. Comparison with text analysis and our proposed sentiment aggregation approach

Approach	Accuracy using Naive Bayes	Accuracy using SVM
Accuracy using only text	0.8765	0.8856
Accuracy using text and emoji both	0.9021	0.910365

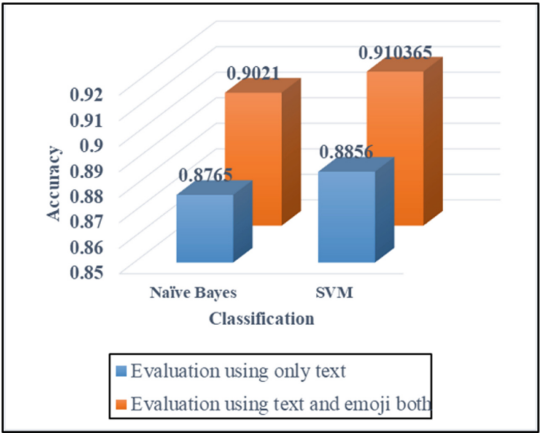


Fig. 5. Comparison with text analysis and our proposed sentiment aggregation approach

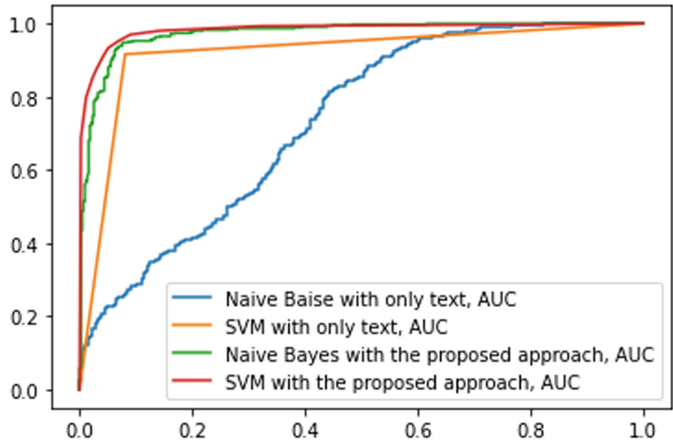


Fig. 6. AUC for model validation

6 Conclusion and Future Scope

Though sentiment analysis is a very familiar topic it is related to the term The metaverse it creates distinct importance. And not only that it is also crucial for automated sentiment analysis technologies to appropriately read these pictorial cues for sentiment as more people employ emoticons in their virtual utterances of thoughts. Our research of the typical function that emoticons do in expressing a text’s overall sentiment and our discussion of how to use this information to our advantage in a lexicon-based sentiment analysis approach constitutes the main contribution of our work. However, in future along with machine learning algorithms, some deep learning algorithms can be applied to the deep emoji concept. Apart from these other relevant datasets can be considered for the performance measurement of the proposed algorithm.

References

1. Arali, Ö., Aydin, Ö.: Tweet classification and sentiment analysis on the metaverse related messages. *J. metaverse* **1**(1), 25–30 (2021)
2. Hayawi, K., Shahriar, S., Serhani, M.A., Alothali, E.: Inevitable the metaverse: a novel twitter dataset for public sentiments on the metaverse (2022)
3. Akkuzukaya, G.: Sentiment analysis on the the metaverse: twitter data. *Sakarya Univ. J. Comput. Inf. Sci.* **5**(2), 147–156 (2022)
4. LeCompte, T., Chen, J.: Sentiment analysis of tweets including emoji data. In: 2017 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 793–798 (2017)
5. Young, M.K., Rieser, J.J., Bodenheimer, B.: Dyadic interactions with avatars in immersive virtual environments: high fiving. In: Proceedings of the ACM SIGGRAPH Symposium on Applied Perception, pp. 119–126 (2015)
6. Wei, Y., Qin, X., Tan, X., Yu, X., Sun, B., Zhu, X.: The design of a visual tool for the quick customization of virtual characters in OSSL. In: 2015 International Conference on Cyberworlds (CW), pp. 314–320. IEEE (2015)
7. Duan, H., Li, J., Fan, S., Lin, Z., Wu, X., Cai, W.: The metaverse for social good: a university campus prototype. In: Proceedings of the 29th ACM International Conference on Multimedia, pp. 153–161 (2021)
8. Chaturvedi, I., Cambria, E., Welsch, R.E., Herrera, F.: Distinguishing between facts and opinions for sentiment analysis: survey and challenges. *Inf. Fusion* **44**, 65–77 (2018)
9. Ullah, M.A., Marium, S.M., Begum, S.A., Dipa, N.S.: An algorithm and method for sentiment analysis using the text and emoticon. *ICT Express* **6**(4), 357–360 (2020)
10. Roy, S., Patel, B., Bhattacharyya, D., Dhayal, K., Kim, T.-H., Mittal, M.: Demographical gender prediction of twitter users using big data analytics: an application of decision marketing. *Int. J. Reason. Based Intell. Syst.* **13**(2), 41–49 (2021)
11. Geetika Vashisht, M.J.: Emoticons emojis based sentiment analysis: the last two decades! *Int. J. Sci. Technol. Res.* **9**(3) (2020)
12. Atif, M., Franzoni, V.: Tell me more: automating emojis classification for better accessibility and emotional context recognition. *Future Internet* **14**(5), 142 (2022)
13. Liu, C., et al.: Improving sentiment analysis accuracy with emoji embedding. *J. Saf. Sci. Resil.* **2**(4), 246–252 (2021)



A Personalized Recommender System Based-on Knowledge Graph Embeddings

Ngoc Luyen Le^{1,2}(✉), Marie-Hélène Abel¹, and Philippe Gouspillou²

¹ Université de Technologie de Compiègne, CNRS, Heudiasyc (Heuristics and
Diagnosis of Complex Systems), CS 60319, 60203 Compiègne Cedex, France
{ngoc-luyen.le,marie-helene.abel}@hds.utc.fr

² Vivocaz, 8 B Rue de la Gare, 02200 Mercin-et-Vaux, France
p.gouspillou@vivocaz.fr

Abstract. Knowledge graphs have proven to be effective for modeling entities and their relationships through the use of ontologies. The recent emergence in interest for using knowledge graphs as a form of information modeling has led to their increased adoption in recommender systems. By incorporating users and items into the knowledge graph, these systems can better capture the implicit connections between them and provide more accurate recommendations. In this paper, we investigate and propose the construction of a personalized recommender system via knowledge graphs embedding applied to the vehicle purchase/sale domain. The results of our experimentation demonstrate the efficacy of the proposed method in providing relevant recommendations that are consistent with individual users.

Keywords: Knowledge graph · Node embedding · Deep learning · Recommender system

1 Introduction

In today's world, where there is an abundance of information, recommender systems have become increasingly important in curating and presenting relevant content to users. These systems are designed to assist individuals by providing personalized suggestions and recommendations, which makes information discovery and decision making more efficient. The use of knowledge graphs for recommender system has demonstrated the ability to produce highly accurate recommendations that are also straightforward to interpret and explain [14, 24]. The essential elements of knowledge graphs are to enhance the organization and structure of information which is able to effectively define a measure of relatedness between entities [6].

A knowledge graph by means of ontology creates a structured framework for a set of concepts or terms within a specific domain by arranging them in a hierarchical manner, and by using relation descriptors to model the connections between these concepts or terms. This provides a standardized lexicon for representing entities within that domain [11, 16]. Recently, the application of knowledge graphs has grown in popularity within the field of recommender systems and decision support systems for graph-based feature learning [20, 25]. As a part of our work, we are

especially interested in constructing a personalized recommender system by utilizing knowledge graph embedding for feature learning.

Most of the work in knowledge graph embedding focuses on create continuous vector representation represented the entities and relations and define a scoring function to measure its relatedness [21]. These vectors are learned using feature learning algorithms such as TransE [1], TransR [22], Node2Vec [4] and others which are designed to capture the structural relationships between entities in the graph. The main idea of knowledge graph embeddings is to map the symbolic entities and relations in the knowledge graph to continuous vector space, this allow us to apply mathematical operations to them, like similarity measure or clustering.

In the field of purchasing and selling vehicles, the descriptions of the vehicles often contain a wealth of information. This information can be organized using ontologies, as demonstrated in works such as [5, 12]. To the best of our knowledge, there does not currently exist a recommender system for the vehicle domain that utilizes features from knowledge graph embeddings.

The rest of this article is organized as follows: Sect. 2 introduces works from the literature on which our approach is based. Section 3 presents our main contributions, starting with the problem statement and then our proposed approach for the construction of the recommender system using knowledge graph embeddings. In Sect. 4, we test our work using a dataset from the domain of purchase/sale of used vehicles. Finally, we conclude and present the perspectives.

2 Related Works

Recommender systems are a type of application that attempts to predict a user's preferences for items and suggests the most relevant items to them through information retrieval [10]. These suggestions can aid users in various decision-making processes, such as choosing music to listen to or products to purchase. Generally, recommender systems are classified into six main categories: Collaborative Filtering, Content-based, Demographic-based, Knowledge-based, Context-aware, and Hybrid [5].

Collaborative Filtering RSs are based on the idea that users who have similar tastes in items will have similar preferences in the future. These systems analyze past interactions between users and items, such as purchase history or ratings, to identify patterns and make recommendations [18]. While content-based RSs, on the other hand, focus on the characteristics of the items themselves to make recommendations. These systems analyze the attributes of items and compare them to a user's past interactions or preferences to make suggestions [9]. With these RS types, the integration of external knowledge resource can provide more information about context or personalize on users and items which particularly useful for improve the accuracy and performance of RSs.

In the past couple of years, there have been many studies that have incorporated knowledge graphs into the construction of feature learning for recommender systems. For example, Sun et al. proposed a novel neural network to capture the semantic information of entity pairs in their work [19]. In [23], the authors presents an intriguing illustration of how knowledge graph embeddings can be integrated with other representations constructed from multimedia content to enhance the

quality of recommendations. They employ translational models to develop knowledge graph embeddings, which are subsequently combined with embeddings of the items' content, such as textual and visual knowledge, to initiate a matrix factorization. In [17], the authors proposed the construction of knowledge embeddings for content-based recommender systems by learning representations of entities in RDF graphs using external knowledge graphs from DBpedia and Wikidata. In [14], the authors introduce a novel approach called *entity2rec* for constructing knowledge graph embeddings based on the properties of entities. They define a global user-item score to compute the relatedness of properties and use a learning ranking algorithm to rank the top recommended items. The above studies have demonstrated that incorporating recommendation algorithms based on knowledge graph embeddings represents a simple and efficient approach for the inclusion of user and item knowledge entities in the recommendation process [7].

In the context of recommender systems, feature learning can be used to automatically discover and extract features from knowledge graph in order to create knowledge graph embedding. These features can be used as input to other algorithms, such as matrix factorization or neural networks. By learning a compact and informative representation of the entities and relationship in knowledge graphs, feature learning can help to capture the most important and relevant information for a recommendation task, leading to more accurate and personalized recommendations for users.

3 Our Proposition

In this section we lay out our proposition for constructing a recommender system based on knowledge graph embedding. For demonstration purposes, we introduce vehicle knowledge graph represented by means of ontologies from the e-commerce domain for the purchase/sale of vehicles.

3.1 Problem Statement

In the context of an e-commerce applications, a typical recommender system consists of three main components: a set of items $I = \{i_1, i_2, \dots, i_n\}$, a set of users $U = \{u_1, u_2, \dots, u_3\}$ and a set of interactions between users and items $D = \{d_{ui} | u \in U, i \in I\}$, where $d_{ui} = 1$ denotes that there are an interaction between user u and item i , $d_{ui} = 0$ means otherwise.

A knowledge graph is a structured representation of entities and their relationships, which can be formalized as a set $G = (E, R)$. The set E comprises of the entities and R denotes the set of relations between these entities. In the context of a recommender system, the entities E include users $u \in U$ and items $i \in I$. The relations between entities are represented as a triple (e, r, e) , where $r \in R$ and $e \in E$. In general the relation can be described by a property of the entity. For example, the transmission of the vehicle item "Tesla Model S 2020" is automatic. This information can be represented in the form of a triple $(Tesla_Model_S, Transmission, Automatic)$. The knowledge graph can be created and structured using a domain ontology or by extracting information from large knowledge bases such as DBpedia or Wikidata.

Given a collection of interactions between users and items in a e-commerce application and the knowledge graph G , the recommendation task is to predict and rank a list of the most suitable items for a particular user from a list of candidate items through compare the score between them.

By utilizing the knowledge graph, we can integrate it into the recommendation task. In the following section, we will investigate a method to extract information from the knowledge graph and use it to enhance the performance of the recommender system.

3.2 Knowledge Graph Embedding Based on Relation Types

The wealth of information provided by the knowledge graph can be leveraged to add additional information to the learning process. Building on the work of Palumbo et al. in *entity2rec* [14, 15], we propose utilizing knowledge graph embeddings by incorporating feature learning for specific properties of entities or relation types from a knowledge graph perspective. Different methods have been developed for feature learning from knowledge graphs to create knowledge graph embeddings, such as TransE [1] or TransR [22]. These methods learn information about entities without considering their specific relation types or properties, which can make the results difficult to understand and explain. In contrast, the *entity2rec* approach [14, 15] considers the relation types and properties used to create the vector representations, allowing for observation and evaluation of the importance of different relation types in knowledge graphs for recommendation tasks.

The vector representation of two nodes should capture as much information as possible about the nodes, including their relation types with other nodes. This is because different relation types have different semantic values based on their relationship with other entities in the knowledge graph. By separating feature learning based on relation types, it is possible to assign a weight to each relation type, which can be used for overall comparison or for comparison of specific properties. For example, if comparing two vehicle entities, they may have the same vehicle type but different manufacturers. In this case, the relatedness based on the “*vehicle type*” relation type will differ when compared to the relatedness based on all relation types.

We adopt the widely used *node2vec* on the knowledge graph to each relation type in order to embed entities into space $\mathbb{R}^d$. First, we extract and build a sub-graph for each relation type G_p . Then, we start to learning vector representation of nodes with each relation type $x_p : e \in G_p \rightarrow \mathbb{R}^d$. The model is trained by using the *node2vec* objective function [4]:

$$\max_{x_p} \sum_{e \in G_p} (-\log Z_e + \sum_{n_i \in N(e)} x_p(n_i) \times x_p(e)) \quad (1)$$

where $Z_e = \sum_{v \in G_p} \exp(x_p(e) \times x_p(v))$ denotes the per-node partition function. We approximate it by using negative sampling [13]. The neighborhood of an entity e , denoted as $N(e)$, is defined using a *node2vec* random walk. The optimization process for this is performed using stochastic gradient ascent on the parameters that define x_p . With the vector representation, $x_p(e)$, for an entity

on a relation type p , we will investigate the method of computing the relatedness between entities in the next section.

3.3 Top-n Recommendations Based on Learning to Rank

Learning to Rank (LTR) can be applied in top-n recommendation tasks to sort a list of items that a user may be interested in. The model is trained using data such as ratings, rankings, or implicit feedback, and it learns to rank items based on this information. Once the model is trained, it can be used to predict the relevance of new items to a given user, and the top-n most relevant items can be recommended to the user [8]. One of the main advantages of using LTR is that it allows for a more fine-grained control over the ranking process. Unlike traditional collaborative filtering methods which tend to recommend items that are similar to items, LTR methods can take into account multiple aspects of the items, such as their content and context, to generate more accurate and personalized recommendations.

Top-n item recommendation for a particular user $u \in U$ is extracted from the n -highest scores from the list of candidate items. Therefore, a ranking function $\tau(u, i)$ is a function that evaluates the relevance of an item to a user by assigning it a score which represents the level of relevance of the item to the user, with higher scores indicating that the item is more relevant. The ranking function's objective is to order a list of items by relevance to the user, with the items having the highest scores considered to be the most relevant. We have a set of sub-graphs extracted from a knowledge graph based on the set of relation types. The entity embedding based on relation types x_p is employed to compute the similarity scores for the ranking function as follows:

$$\tau(u, i) = \mathfrak{S}(x_p(u), x_p(i)) \quad (2)$$

where $\mathfrak{S}(x_p(u), x_p(i))$ denote the cosine similarity between item vector $x_p(i)$ on relation type p and user vector $x_p(u)$ on relation type p .

In our work, we employ LambdaMART [2] which learn to ranking based on the gradient boosting framework. It is a tree-based algorithm that is designed to optimize the pairwise or list-wise loss functions, which measure the difference between the predicted order of items and their true order. LambdaMART uses a combination of multiple decision trees to make the ranking predictions, and it optimizes the parameters of the trees using gradient descent.

In the next section, we will be introducing our training model that is based on feature learning utilizing relation types present in the knowledge graph. Additionally, a ranking function will be implemented to optimize the model. This approach allows us to effectively learn important features and improve the overall performance of the model.

3.4 Training Model and Optimisation

Given the interaction dataset from the user set U on item set I and the knowledge graph G , each user u_k is associated with a set of items from the interaction data $d_k = \{d_{k1}, d_{k2}, \dots, d_{kn}\}$, and the set of labels $\vec{y}_k = \{y_{k1}, y_{k2}, \dots, y_{kn}\}$, the training

set is defined $X_{ds} = \{(u_1, \vec{d}_1, \vec{d}_1), (u_2, \vec{d}_2, \vec{y}_2), \dots, (u_N, \vec{d}_N, \vec{y}_N)\}$, where N is the number interaction between users and items in the dataset.

With the ranking function $\mathbf{r}(u, i)$, we can compute the relatedness score $\vec{\mathbf{r}}(u, i)$ for pairs of user u and item i . The top- n recommendation task becomes learning a prediction function $y_k = f(\mathbf{r}(u, i), \theta)$. with the loss function is computed as follows [14]:

$$L(\theta) = \sum_{k=1}^N (1 - \mathcal{A}(\pi(u_k, \vec{d}_k, \theta))) \quad (3)$$

where $\mathcal{A}(\pi(u_k, \vec{d}_k, \theta))$ denote the agreement between the permutation $\pi(u_k, \vec{d}_k, \theta)$ induced from $\mathbf{r}(u_k, \vec{d}_k, \theta)$ and the ground truth relevance of items. Therefore, the goal of the learning process is to find the set of parameters θ that minimize the loss function $L(\theta)$ on the training data X_{ds} .

4 Experiments

In this section, we evaluate our approach on a real case of vehicle purchase/sale domain. The structure and organization of the vehicle's description and user profiles are carried on by using ontologies that can queried by using SPARQL language [12].

4.1 Dataset and Evaluation

We extract and built our dataset from vehicle purchase/sale domain that include users are clients and items present for vehicle models on the market. We summarize the statistic of our datasets in Table 1.

Table 1. Some statistics on the dataset

User - Item interaction	Users	393
	Items	5.537
	Iterations	99.121
Knowledge graph	Entities	27.561
	Relation types	6
	Triplets	822.000

In order to evaluation our work, we use a standard information retrieval metrics such as Precision at n , recall at n and Mean Average Precision (MAP) that can be used to evaluate the performance of a recommender system. Precision @ n shows the proportion of recommended items that are relevant to the user among the top- n recommendations and is defined as follows:

$$P(n) = \frac{1}{|U|} \sum_{u \in U} \sum_{k=1}^n \frac{can(i_k, u)}{n} \quad (4)$$

where $can(i_k, u) = 1$ if the candidate item i is relevant to user u , otherwise $can(i_k, u) = 0$. While recall @k is the proportion of relevant items that are recommended among all relevant items and is calculated as follows:

$$R(n) = \frac{1}{|U|} \sum_{u \in U} \sum_{k=1}^n \frac{can(i_k, u)}{rel(u)} \quad (5)$$

where $rel(u)$ is the relevant items set for user u in the test set. The third metric MAP allows a measure of the average precision at n for a set of item recommendations.

$$MAP(n) = \frac{1}{|U|} \sum_{u \in U} \sum_{k=1}^n \frac{can(i_k, u)}{n} \times rel(i_k, u) \quad (6)$$

These metrics can be used to assess the accuracy of the recommendations made by the recommender system, with higher values indicating better performance.

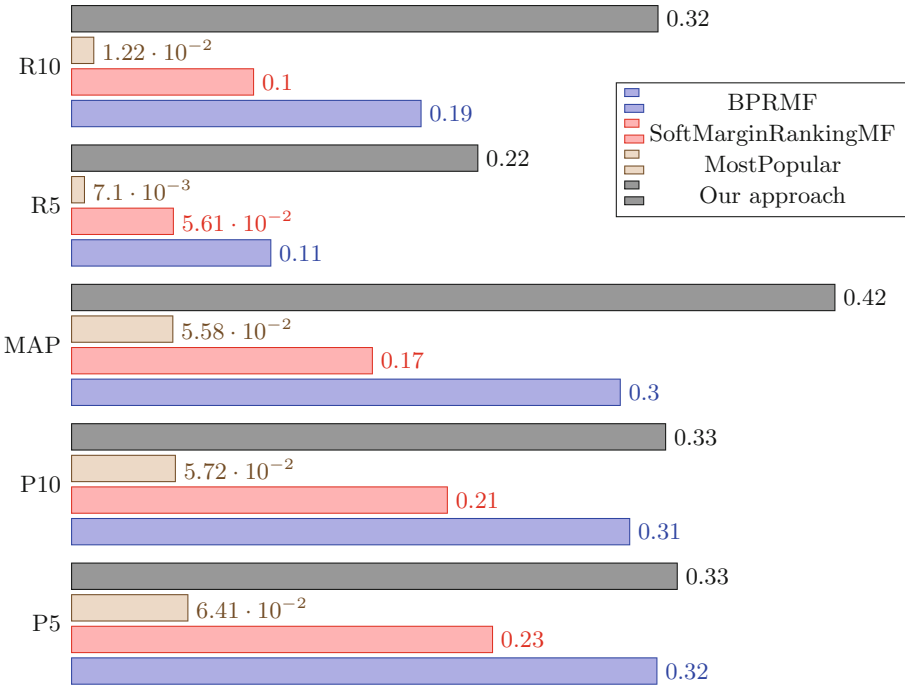


Fig. 1. Comparison with other approaches on the dataset

4.2 Results

To conduct a thorough experiment on the dataset, we have pre-configured certain hyper-parameters on the model to optimize its performance. One of the key hyper-parameters that we have set is the dimension of embeddings, which we have set to 200. This value was chosen based on prior research and experimentation, and it has been found to be effective in capturing the underlying structure of the data. Another important hyper-parameter that we have set is the number of random walks per node in the graph. This value has been set to 100 and is used to explore the connectivity of the nodes in the graph. In addition to these specific hyper-parameter settings, we have also retained the default values in the LambdaMart learning to rank and node2vec models.

The experiments focus on extracting and comparing knowledge graph embeddings of different features. Specifically, the feature sets are defined as a set of six features $\{Vehicle\ style, Fuel\ style, Mileage, Number\ of\ seats, Transmission\ type, Vehicle\ price\}$. These features have been chosen as they are considered to be the most relevant and informative attributes of a item vehicle, and they are commonly used in the automotive domain to describe and compare different models. The goal of the experiments is to evaluate these features by supplementing them with information from knowledge graphs using relation type embeddings for each item vehicle.

From the experimental results depicted in Fig. 1 and Table 2, several key insights can be gleaned. Our proposed approach demonstrates superior performance in terms of top-n recommendations across various evaluation metrics when compared to other approaches [3] such as MostPopular, SoftMarginRankingMF, and BPRMF on the vehicle dataset. This serves as evidence that incorporating features from a knowledge graph can provide additional information that improves the prediction of top-n items for a given user. Additionally, it highlights the potential of using knowledge graph-based features to enhance recommendation systems.

Table 2. Result of on different feature evaluation from the dataset

Relation type feature	P5	P10	MAP	R5	R10
Vehicle style feature	0.3491	0.3501	0.4509	0.2179	0.3268
Fuel type feature	0.3348	0.3412	0.4295	0.2113	0.3163
Mileage feature	0.2534	0.2811	0.3890	0.1889	0.2968
Number of seats feature	0.3160	0.3017	0.4190	0.2051	0.2897
Transmission type feature	0.3791	0.3567	0.4311	0.2228	0.3296
Vehicle price feature	0.3557	0.3374	0.4059	0.2132	0.3057

In this work, we aimed to investigate the role of different features extracted from a knowledge graph through relation type embeddings. The results presented in Table 2 demonstrate that there are variations in the achieved results

when utilizing different features. By separating features by relation type, we can gain insight into the factors that influence item recommendations for users. As a result, we used these embeddings to evaluate the importance of individual features in relation to the overall results obtained. Additionally, it allows us to identify which relation types are more informative for recommendation and which are less important, and use it to propose an interpretation to users, which can make the recommendation system more transparent, and can also help to improve the user's understanding and trust of the system.

5 Conclusion and Perspectives

In this paper, we investigate the construction of an approach for a personalized recommender system based on building knowledge graph embedding, which allows for the extraction and addition of more information into the learning-to-rank process. From our work, we illustrate how we can separate building sub-graphs for each relation type in the knowledge graph and use it to build embeddings. The results obtained from the experiment show that our approach is of interest as it provides better results compared to other approaches on our dataset. In our future work, we intend to research the exploitation of ways to explain the recommendation results to a given user by analyzing features learned from knowledge graphs.

Acknowledgment. This work was funded by the French Research Agency (ANR) and by the company Vivocaz under the project France Relance - preservation of R&D employment (ANR-21-PRRD-0072-01).

References

1. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. *Adv. Neural Inf. Process. Syst.* **26** (2013)
2. Burges, C.J.C.: From RankNet to LambdaRank to LambdaMART: an overview. *Learning* **11**(23–581), 81 (2010)
3. Gantner, Z., Rendle, S., Freudenthaler, C., Schmidt-Thieme, L.: MyMediaLite: a free recommender system library. In: 5th ACM International Conference on Recommender Systems (RecSys 2011) (2011)
4. Grover, A., Leskovec, J.: node2vec: scalable feature learning for networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–864 (2016)
5. Le, N.L., Abel, M.-H., Gouspillou, P.: Towards an ontology-based recommender system for the vehicle sales area. In: Troiano, L., Vaccaro, A., Kesswani, N., Díaz Rodríguez, I., Brigui, I. (eds.) *ICDLAIR 2021. LNNS*, vol. 441, pp. 126–136. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-98531-8_13
6. Le Ngoc, L., Abel, M. H., Gouspillou, P.: Apport des ontologies pour le calcul de la similarité sémantique au sein d'un système de recommandation. In: *Ingénierie des Connaissances (Evènement affilié à PFIA'22 Plate-Forme Intelligence Artificielle)*, Saint-Étienne, France, June 2022

7. Liu, C., Li, L., Yao, X., Tang, L.: A survey of recommendation algorithms based on knowledge graph embedding. In: 2019 IEEE International Conference on Computer Science and Educational Informatization (CSEI), pp. 168–171. IEEE (2019)
8. Liu, T.-Y., et al.: Learning to rank for information retrieval. *Found. Trends® Inf. Retr.* **3**(3), 225–331 (2009)
9. Lops, P., de Gemmis, M., Semeraro, G.: Content-based recommender systems: state of the art and trends. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) *Recommender Systems Handbook*, pp. 73–105. Springer, Boston, MA (2011). https://doi.org/10.1007/978-0-387-85820-3_3
10. Jie, L., Dianshuang, W., Mao, M., Wang, W., Zhang, G.: Recommender system application developments: a survey. *Decis. Support Syst.* **74**, 12–32 (2015)
11. Luyen, L.N., Tireau, A., Venkatesan, A., Neveu, P., Larmande, P.: Development of a knowledge system for big data: case study to plant phenotyping data. In: *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics, WIMS 2016* (2016)
12. Gouspillou, P., Le, N.L., Abel, M.H.: Improving semantic similarity measure within a recommender system based-on RDF graphs. In: *Proceedings of the 6th International Conference on Information Technology & Systems* (2023)
13. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **26** (2013)
14. Palumbo, E., Monti, D., Rizzo, G., Troncy, R., Baralis, E.: Entity2rec: property-specific knowledge graph embeddings for item recommendation. *Expert Syst. Appl.* **151**, 113235 (2020)
15. EPalumbo, E., Rizzo, G., Troncy, R.: Entity2rec: learning user-item relatedness from knowledge graphs for top-n item recommendation. In: *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 32–36 (2017)
16. Rodriguez, M.A., Egenhofer, M.J.: Determining semantic similarity among entity classes from different ontologies. *IEEE Trans. Knowl. Data Eng.* **15**(2), 442–456 (2003)
17. Rosati, J., Ristoski, P., Di Noia, T., Leone, R.D., Paulheim, H.: RDF graph embeddings for content-based recommender systems. In: *CEUR Workshop Proceedings*, vol. 1673, pp. 23–30. RWTH (2016)
18. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *The Adaptive Web. LNCS*, vol. 4321, pp. 291–324. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72079-9_9
19. Sun, Z., Yang, J., Zhang, J., Bozzon, A., Huang, L.K., Xu, C.: Recurrent knowledge graph embedding for effective recommendation. In: *Proceedings of the 12th ACM Conference on Recommender Systems*, pp. 297–305 (2018)
20. Wang, K., Zhang, T., Tianqiao Xue, Y.L., Na, S.-G.: E-commerce personalized recommendation analysis by deeply-learned clustering. *J. Vis. Commun. Image Represent.* **71**, 102735 (2020)
21. Wang, Q., Mao, Z., Wang, B., Guo, L.: Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans. Knowl. Data Eng.* **29**(12), 2724–2743 (2017)
22. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28 (2014)

23. Zhang, F., Yuan, N.J., Lian, D., Xie, X., Ma, W.Y.: Collaborative knowledge base embedding for recommender systems. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 353–362 (2016)
24. Zhang, Y., Chen, X., et al.: Explainable recommendation: a survey and new perspectives. *Found. Trends® Inf. Retr.* **14**(1), 1–101 (2020)
25. Zhu, Q., Zhou, X., Jia, W., Tan, J., Guo, L.: A knowledge-aware attentional reasoning network for recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 6999–7006 (2020)



Awareness of Arabic Teachers in Academic Integrity Towards Students Completing Tasks in Abu Dhabi: Data Analysis Approach

Suad Abdalkareem Alwaely^{1,2} , Asaad Ali Muslim Al-Maamari³,
and Muhammad Turki Alshurideh^{4,5}

¹ College of Education, Humanities and Social Sciences, Al Ain University, Al Ain, UAE
Suad.alwaely@aau.ac.ae

² Hashemite University, Zarqa, Jordan

³ Ministry of Education, Dubai, United Arab Emirates

⁴ Department of Marketing, School of Business, The University of Jordan, Amman 11942, Jordan

m.alshurideh@ju.edu.jo, malshurideh@sharjah.ac.ae

⁵ Department of Management, College of Business, University of Sharjah, Sharjah 27272, United Arab Emirates

Abstract. This paper aims to investigate the level of Arabic language teachers' awareness of the values of academic integrity and its relationship to the motivation to complete tasks among students of the third cycle in the Emirate of Abu Dhabi. The study sample consisted of (50) male and female teachers, and their students; (150) male and female students. The study followed the descriptive analytical approach and adopted two tools: the questionnaire to measure the degree of awareness of Arabic language teachers about the values of academic integrity, and the task completion questionnaire for the twelfth-grade students. The results of the study showed that the level of awareness of Arabic language teachers about the values of academic integrity was high. It was also noted that there were statistically significant differences in the level of awareness of Arabic language teachers for the twelfth grade due to the gender variable and such differences were in favor of males. It was also noted that there were no statistically significant differences due to the educational qualification variable, and there were differences due to the educational qualification (bachelor's degree) variable. The results of the questionnaire's analysis of the motivation to complete the tasks by the twelfth-grade students showed that it was statistically significant. Based on the results, the study came out with several recommendations.

Keywords: Awareness level · Achievement motivation · Academic integrity · The third cycle

1 Introduction

Academic integrity is the moral framework that civil institutions seek to establish to control people's lives and give them access to values that will guide their actions. Without

ethical standards that society upholds, defend individual rights, and serve the public good, democratic life has no significance (Alshurideh et al. 2017; Fadloun 2017; AlShehhi et al. 2020). Furthermore, integrity means serving citizens and gaining their trust, which is the goal of a public office. Therefore, the public employee must strive to maintain and enhance the confidence of citizens, stress the importance of the integrity of government administration and work to enhance the public interest of society, by not using his powers and position in an improper manner, and to prioritize the public interest over his personal interests in case there is a conflict of interests, and to disclose cases of fraud and mismanagement in case they are accessible to him, and not to disclose official information in an incorrect manner or use it for personal purposes (Alsuwaidi et al. 2020; Al-Shukri 2003; Kurdi et al. 2020).

The promotion of integrity values in governmental institutions - including educational institutions - is closely linked to the existence of a sound institutional structure that takes into account these values and principles in its management, through the availability of effective systems of transparency, accountability and participation, in addition to the presence of working staff committed to values based on work for the good, in addition to the existence of effective oversight bodies that enjoy high independence and professionalism (Khan et al. 2017; Al Kurdi et al. 2020).

Integrity necessitates comprehensive reform programs with strong political backing and strategic content based on identifying the issue, resolving its root causes, involving the public and its institutions, establishing and reinforcing moral and ethical standards for government and society, and learning from global experiences. (Alshurideh et al. 2016; Alshurideh et al. 2019; Alyammahi et al. 2020). This is because protecting integrity and combating corruption in all its forms are among the established principles of Islamic Law (Sharia) and international regulations (Dweik 2013).

This study deals with the problem of academic integrity in educational institutions, the most important of which are schools, as they are the incubators of the educational process. Therefore, highlighting the importance of the total quality system and academic accreditation is one of the basic mechanisms for developing a culture of academic integrity and increasing respect and local appreciation and practical recognition of the educational institution for the services it provides to students and the community by contributing to the development of the local community (Alshurideh 2022; Alzoubi et al. 2023). This will only be achieved by developing the values of sincerity, honesty and belonging to the educational institution and eliminating all methods of cheating and scientific theft in education and in scientific research (Alshurideh et al. 2022; Kurdi et al. 2023). Therefore, attention to the culture of academic integrity has become a concern that preoccupies the minds of educators to achieve quality, especially in the sense that we are in the era of the information revolution and the flow of knowledge and the control of the ethics of academic integrity between the teacher, the learner and all workers in society (Alsharari and Alshurideh 2020; Al Shraah et al. 2022). This makes the entire educational process on the right track and raises the value of Educational Justice (Abdul-Saheb 2016).

The researchers believe that integrity is a set of systems, laws, principles and values that teachers must possess, and they are derived from regulations and legislations issued by educational institutions, which regulate the learning process. There may be many

forms of academic integrity violations, such as non-compliance with lecture dates, failure to prepare for it, frequent absences, violating regulations, covering up behaviors in violation of laws, and failure to report transgressors. The following actions and behaviors are the most significant examples of behavior violating academic honesty standards and values (American University of Beirut 2003, Najmuddin 2017, Al Dahshan 2018; Leo et al. 2021; Nuseir et al. 2021).

Cheating: It occurs through using the works of others, unauthorized use, intellectual theft or plagiarism, and it is related to quoting the words and works of others. **Misuse of academic assistance:** This happens through a student's use of the work of his colleague who provides him with academic assistance, fabrication and forgery by the student changing the information available to him or fabricating it in an exam or an academic assignment and impersonating the personality through the student claiming to be someone else in the classroom. It can also take the form of multiple submission of the same work by submitting identical research papers or works for a subject to obtain grades and for more than one time without prior permission from the teacher, and teaching, by not making sure that students understand and achieve the objectives of the lesson or introduce some students to others. This does not apply to the values of integrity and scientific research. It is to take the words, ideas or research of others and put them in your work without referring to him and his work, so that it becomes yours. **Local community service:** This occurs through the provision of educational data and statistics contrary to reality. Motivation is linked to integrity. It is a manifestation of a feeling of dedication to work, striving for efficiency in performing the tasks entrusted to the person, continuous desire for achievement, perseverance, continuing to work for long periods, and overcoming the obstacles facing people to achieve performance and efficiency. And these manifestations are all indicators of achievement motivation, a fundamental dimension for individuals and educated people (Khuwailid 2016a; 2016b).

Balash and Suliha (2018) stated that the motivation for achievement is a basic and important factor in guiding the behavior of the individual and the behavior of others around him, as it is an essential component of the individual's self-realization and assertion, as well as to ensure his continuity in life and the attainment of a better lifestyle. Educational institutions have focused on this concept, which made them provide alternatives and positive returns for learners that enhance their satisfaction and motivation factors and guarantee them the highest level of positive returns (Amarneh et al. 2021; Alshurideh et al. 2023).

Khalifa (2013) believes that the achievement motivation is not limited to merely stimulating behavior but extends to directing this behavior to satisfy the motive, where this sequence ends with the achievement of the goal; that is, the subject of the motive. It becomes clear that every behavior, whatever its form, has a set of motives behind it, directing it towards the search to achieve the satisfaction associated with this motive. Moreover, motives have the power to move behavior and direct it in a certain direction towards restoring balance. The individual seeks to achieve the goal with a degree of excellence and distinction.

Despite the early beginnings of the study of achievement motivation, the credit goes to the psychologist Henry Murray who was the first scholar to present the concept of

achievement motivation precisely. Murray considered achievement motivation a component of personality. He is also credited with starting to define the concept of this motivation and establishing the rules that can be used to measure it. Motivation is provoked either by subjective internal factors (needs, tendencies, and interests) or external environmental factors such as (people, ideas, and things) (Bettayeb et al. 2020; Al-bawaia et al. 2022; Al-kharabsheh et al. 2023). For this reason, teachers who wish to raise the level of internal motives of their students towards exerting academic effort must learn how to prepare the study environment for their advancement (Spaulding 1992).

2 Previous Studies

Khan et al. (2017): The purpose of the study was to shed light on how Jordanian universities promote integrity principles and how that relates to a sense of national identity among both private and public university students. Its results showed that the role of the Jordanian universities in promoting integrity values and in the sub-fields: transparency, accountability, and participation showed medium averages, and the level of national affiliation came in high averages, and there were differences in promoting integrity values attributed to the type of university and in favor of private universities.

Akram's Study (2017): studied the effect of using the Telegram Channel in enhancing the integrity values and attitudes of forensic science female teachers. The study found statistically significant differences in the mean scores of the variables in the pre and post application of the values scale and the orientation scale, indicating that using the Telegram channel has an effect on improving the values of integrity and the orientation towards it.

Hamid and Kharbit Study (2018): It aimed to find out the effect of teaching language functions on enhancing learners' motivation and fluency. Achievement in language functions was measured for third-year students in the Department of English, College of Education for Human Sciences, University of Basra. The study found results, the most important of which is the effect of teaching language functions in enhancing motivation and completing tasks.

Saif (2014): examined the relationship between self-confidence, motivation, and academic achievement among students at the College of Education at Taiz University in Yemen. Statistically significant correlations were found between self-confidence, intrinsic motivation, external motivation and achievement, and between achievement and intrinsic motivation and extrinsic motivation.

3 Problem Statement

Teachers' practice of various forms of academic dishonesty impedes the improvement of the academic skills of their students, in addition to weakening the effectiveness of their strategies for memorizing, recalling, and employing information. Accordingly, academic integrity in educational institutions is a situational behavior that is influenced by psychological, behavioral, cognitive, and social variables, which are clearly reflected in the motivation of their students. In addition, the educational system in the educational institution did not address the issue of academic integrity among teachers and the impact

of this on students' motivation to complete tasks. This is an issue that haunts decision-makers and education officials. Therefore, this study came to address the level of Arabic language teachers' awareness of the values of academic integrity and its relationship to the motivation to complete tasks among their students in the Emirate of Abu Dhabi in the United Arab Emirates. This was confirmed by (Al-Harbi 2016; Badrakhn et al. 2017). The problem of the study is manifested in the following questions:

1. What is the level of awareness of 12th-grade Arabic language teachers about the values of academic integrity?
2. Does the level of 12th-grade Arabic language teachers' awareness of the values of academic integrity differ due to (gender, educational qualification, and teaching experience)?
3. Is there a correlation between the level of Arabic language teachers' awareness of the twelfth-grade values of academic integrity and the motivation to accomplish tasks among their students?

4 Purpose and Significance of the Study

The study aimed to find out "the level of Arabic language teachers' awareness of the values of academic integrity and its relationship to the motivation to complete tasks among students of the third cycle in the Emirate of Abu Dhabi".

4.1 Significance of the Study

1. The current study's theoretical significance comes from its emphasis on the academic integrity variable, a relatively new variable in Arab and Emirati educational circles because it is a moral, educational, civilizational, and legal objective in school learning environments. Academic integrity is also a legal and religious requirement. In addition, its significance is partly because improving students' knowledge of and motivation for achievement is one of the critical elements in making the educational process successful.
2. The practical significance of the current study is that educational leaders attempt to capitalize on its findings by supplying data on the study's variables so they can develop educational programs and use educational, educational, and psychological legislation and methods to assist students in subsequently increasing their understanding and awareness of academic integrity standards. This is accomplished by energizing education, research, and volunteer work.

4.2 Study Limits

- Time: The study was conducted during the academic year's second semester (2019–2020).
- Place: This study was limited to a sample of the third-cycle schools (males and females) in the Emirate of Abu Dhabi.

- **Topic:** This study was limited to the level of awareness of Arabic language teachers about the values of academic integrity and the level of their student's understanding of the motivation to complete tasks in the third cycle in the Emirate of Abu Dhabi.
- **Sample:** Arabic language teachers (males and females), who teach the twelfth grade in the Emirate of Abu Dhabi schools.
- The study is determined by the validity and reliability of its tools.

4.3 Procedural Definitions

- **The level of awareness:** It is the understanding based on sensation and knowledge that helps to make certain decisions about a particular issue. In the current study, awareness was measured through the tool prepared by the researchers.
- **Academic Integrity Values:** It is a set of systems, laws, and principles that teachers must possess. It is measured procedurally by the total scores obtained by the examinee (the teacher) as a result of his answer to the Academic Integrity Survey in its three areas (teaching, scientific research, and community service).
- **Achievement Motivation:** It is the tendency of twelfth grade students to reach a specific goal, whether inside or outside the school. It is procedurally measured by the total scores obtained by the examinee (the student) as a result of his answer to the achievement motivation questionnaire, which consists of twenty items.
- **The third cycle:** The researchers define it procedurally within the limits of its study as the twelfth grade, which is the last grade of the third cycle classes in general education in the Ministry of Education in the United Arab Emirates.

5 Methodology

The study followed the descriptive-analytical approach, which is based on diagnosing the current situation by describing the degree of awareness of Arabic language teachers about the values of academic integrity and its relationship to the motivation to complete tasks among students of the third cycle in the Emirate of Abu Dhabi and their responses to the two study tools: the academic integrity questionnaire and the motivation to complete tasks questionnaire.

5.1 Study Sample and Population

The population in this study comprise all the Arabic language teachers ($n=230$) who teach the third cycle (the secondary stage) in the Emirate of Abu Dhabi. The students' number was (5750) male and female students enrolled for the academic year 2019 /2020. The researchers randomly chosen 25% of the male and female teachers' population. Moreover, they excluded a pilot sample ($n=30$) of teachers and determined the study tool's validity coefficient. Thus, the number of the study sample was (50) teachers; (29) male teachers and (21) female teachers. The researchers also selected (3) students taught by each teacher; thus, the study sample consisted of (150) male and female students.

5.2 Study Tools

The first tool: (Academic Integrity Values Questionnaire): It consisted of three areas: Academic Integrity values related to teaching (14 paragraphs), Academic Integrity values related to scientific research (11 paragraphs), and the academic values of the second tool (11 paragraphs). The second tool: (Achievement Motivation Questionnaire): It included (20) paragraphs specifying the motivation to complete tasks for male and female twelfth grade students.

5.3 The Validity and Reliability of the Two Study Tools

To ensure the validity of the study instrument, they were presented to the panel of arbitrators, including professors of curricula and teaching methods, professors of educational psychology in various Emirati universities, and teachers and supervisors of the Arabic language in the Ministry of Education. Moreover, to verify the reliability of the study tools, the test-retest method was used, by applying them to a random exploratory sample consisting of (30) teachers and (30) students twice, with a time difference of two weeks. Using the Pearson Correlation coefficient between the two applications, the correlation coefficient was obtained..

The correlation coefficient between the two applications on the teachers' tool was (0.85), and the correlation coefficient between the two applications on the students' tool was (0.87), which are statistically significant values at the significance level ($\alpha \leq 0.05$).

5.4 Variables

Independent Variables

1. Gender, and has two categories: (male and female).
2. Academic qualification, and has two levels: (Bachelor's, Postgraduate).
3. Teaching experience, and it has three levels (less than 5 years, 5–10 years, and more than 10 years).

Dependent Variables

1. Arabic language teachers level of awareness about the values of academic integrity.
2. Motivation to accomplish tasks among their students in the twelfth grade.

Statistical Methods

- Pearson correlation coefficient was used to determine the correlation between the level of Arabic language teachers' awareness of the values of academic integrity and its relationship to achievement motivation among twelfth-grade students.

1. Frequencies and percentages are used to distribute participants according to personal variables.
2. Arithmetic means and standard deviations: They are used to identify the level of awareness of Arabic language teachers about the values of academic integrity and its correlation to achievement motivation among twelfth-grade students.
3. ANOVA 3 way: It was employed to find the differences in the responses of the sample members about the level of their awareness of the values of academic integrity according to (gender, educational qualification, and teaching experience).
4. Scheffe's test for dimensional comparisons: it is used to identify the differences in the responses of the sample members about the level of their awareness of the values of academic integrity according to the variable of (teaching experience).

6 Results and Discussions

First: Question One: What is the level of awareness of the Arabic language teachers of the twelfth grade about the values of academic integrity?

To answer this question, the means and standard deviations were extracted. It aims to identify the level of awareness of Arabic language teachers about the values of academic integrity and its relationship to the motivation to complete tasks by students of the twelfth grade in the schools of Abu Dhabi. Table 1 illustrates this.

Table 1. Arithmetic means and standard deviations of the answers of the members of the sample of teachers about the fields of the study tool and the tool as a whole (n = 50)

Rank	No	Field	Arithmetic mean	Standard Deviation	Level of awareness
1	1	Teaching	4.48	0.37	High
2	3	Local community service	4.43	0.37	High
3	2	Scientific research	4.27	0.41	High
Total			4.40	0.31	High

Table 1 reflects that the means of the answers of participants to the areas of the study tool ranged between (4.27–4.48), with a high level of awareness for all areas. The field of “teaching” ranked first with a mean of (4.48), and the field of “local community service” ranked second with a mean of (4.43), while in the third and last place came the field of “scientific research” with an average of (4.27). The overall mean of the tool was (4.40) with a high level of awareness. This indicates that the level of awareness of Arabic language teachers in the twelfth grade about the values of academic integrity was high.

The researchers explain this result that Arabic language teachers realize that achieving the values of academic integrity in any society cannot be achieved without the educational sector practicing its principles and values, as the education sector is vital in any society due to its direct relationship with the most important segment of society (students). Teachers also have an important role in educating young people on the

importance of applying the values of integrity in all dimensions and aspects of life. The two researchers discussed the results of each field of the academic integrity values questionnaire, according to the rank they obtained individually, as follows:

The results related to the field of teaching reflected that the level of awareness of Arabic language teachers about the values of integrity in the field of (teaching) was high. The results showed that the Arabic language teachers are interested in adhering to the timetables of the lessons, and they seek by themselves to develop teaching strategies and methods in a way that enhances the values of integrity and refrain from forcing their students to purchase scientific books of their own or those of people they know. And the items related to these values got the highest arithmetic means compared to other items.

This result is attributed the fact that teachers deal with the school as an educational and social system in which teachers and students participate. In this system, a set of social relations, that affect the formulation of individuals' thought and culture, arise. This result can be justified by the decisive role of the teacher in developing behavioral values, shaping the student's identity and increasing his motivation to accomplish the tasks entrusted to him. This result agreed with the result of (Al-Shanqiti 2015).

The researchers explain the result related to the field of community service that Arabic language teachers seek to develop the students' role in community service, enhance their leadership skills and help them exchange ideas by urging them to participate in seminars and competitions to demonstrate their ability to excel and be creative in that, as well as the teachers' interest in students' exercise of freedom to express their opinions and prove their own identity and instill the spirit of leadership in themselves. This result can be justified by the positive role of teachers in building a conscious generation to serve the local community. And this result agreed with (Sulaiman 2015).

The researchers explain the result of the field of practical research through the awareness of Arabic language teachers towards the importance of participating in local and regional conferences related to integrity and transparency for the purposes of developing their academic performance, especially in light of the cultural globalization and the dissolution of values that society is exposed to, which makes teachers try to determine the integrity values that must strengthen and support it, in addition to their relentless endeavor to preserve information channels from (old sources and modern references) that confirm the depth of commitment to cultural identity. This constitutes an important element of the educational personality that the educational institution seeks to strengthen. And this result agreed with the (Khan et al. 2017).

Second: Question Two: does the Level of Arabic language Teachers' Awareness of the Twelfth-Grade Values of Academic Integrity Differ According to (Gender, Educational Qualification, and Teaching Experience)?

To answer this question, averages and standard deviations were extracted, which measures the level of awareness of Arabic language teachers in it according to personal variables (gender, educational qualification, and teaching experience). The 3-way ANOVA was also utilized to the overall academic integrity values according to (gender, educational qualification, and teaching experience). Tables with numbers (2–6) illustrate this.

Table 2. Arithmetic means and standard deviations of participants’ responses by (gender, educational qualification, and teaching experience)

Variable	Level	Arithmetic mean	Standard deviation
Gender	Male	4.31	0.13
	Female	4.05	0.21
Academic qualification	Bachelor’s	4.16	0.23
	Higher education	4.18	0.20
Teaching experience	Less than one year	4.08	0.22
	5–10 years	4.25	0.16
	More than ten years	4.20	0.27

Table 2 shows apparent differences between the mean of the participants’ responses on the academic integrity values by (gender, academic qualification, and teaching experience). And to find out the statistical significance of these differences, a three-way ANOVA was applied. Table 3 illustrates this.

Table 3. The results of the 3-way ANOVA by (gender, educational qualification, and teaching experience)

Variable	Total squares	Freedom levels	Arithmetic mean (Khalil Al-Rashedi)	F	Statistical significance
Gender	1.70	1	1.70	55.73	0.00
Academic qualification	0.07	1	0.07	2.23	0.14
Teaching experience	0.31	2	0.15	4.99	0.01
Error	3.52	45	0.03		
Total	2088.80	50			
As	5.90	49			

Table 3 reflects that:

1. Statistically significant differences at the significance level ($\alpha \leq 0.05$) due to the gender variable. In this respect, the value of (F) reached (55.73), which is a statistically significant value, and by referring to the arithmetic means shown in Table 8, it is clear that the differences are in favor of males with an arithmetic mean of (4.31).
2. No statistically significant differences at the significance level ($\alpha \leq 0.05$) between the means of the participants’ estimates of the academic integrity values due to

the educational qualification variable. The value of (F) was (2.23), which is not a statistically significant value.

3. There are statistically significant differences at the significance level ($\alpha \leq 0.05$) between the means of the participants' estimates on the academic integrity values as a whole, according to the experience variable. The value of (F) reached (4.99), which is a statistically significant value. To find out the sources of these differences, the Scheffe test was utilized for obtaining dimensional comparisons. Table 4 illustrates this.

Table 4. The results of the Scheffe test for post-comparisons

Teaching experience	Arithmetic mean	Less than 5 years	5-10 years	More than 10 years
Less than 5 years	4.08		*0.17-	*0.12-
5-10 years	4.25			0.05
More than 10 years	4.20			

With an arithmetic mean of more than 10, Table 4 shows that the sources of differences were between experience levels (less than 5 years and more than 10 years), favoring the level of experience (more than 10 years) (4.20). The outcomes also demonstrated that there are differences between experience levels (less than five years and five to ten years) that favor the experience level (10 to five years), with an average of (4.25).

The means and standard deviations of the participant's estimates on each domain of academic integrity values were computed by (gender, academic qualification, and teaching experience). Also, 3-way MANOVA was applied to the domains of academic integrity values according to the variables (gender, educational qualification, and teaching experience). Tables 5, 6 illustrate this.

Table 5 shows apparent differences between the means of the participants' responses on the areas of academic integrity values by (gender, academic qualification, and teaching experience). In order to find out the statistical significance of these differences, a 3-way MANOVA was applied. Table 6 shows this.

Table 6 reflects that:

1. No statistically significant differences at the significance level ($\alpha \leq 0.05$) between the overall means of the participants' estimates on the field of (teaching) were evident due to the gender variable. Besides, the value of (F) reached (3.34), which is not a statistically significant value.
2. Statistically significant differences at the significance level ($\alpha \leq 0.05$) were found on the areas (scientific research and local community service) according to the gender variable. In this respect, the values of (F) were statistically significant. By reference

Table 5. Arithmetic means and standard deviations of the participants' responses by (gender, educational qualification, and teaching experience)

Variable	Level	Teaching		Scientific research		Local community service	
		Arithmetic mean	Standard deviation	Arithmetic mean	Standard deviation	Arithmetic mean	Standard deviation
Gender	Male	4.28	0.24	4.28	0.28	4.38	0.27
	Female	4.17	0.31	4.11	0.28	4.01	0.37
Academic qualification	Bachelor's	4.23	0.28	4.20	0.30	4.15	0.38
	Postgraduate	4.13	0.31	4.13	0.25	4.30	0.32
Teaching experience	Less than 5 years	4.15	0.29	4.12	0.28	4.04	0.38
	5–10 years	4.28	0.23	4.25	0.28	4.32	0.30
	More than 10 years	4.24	0.34	4.20	0.33	4.20	0.39

to the arithmetic means shown in the previous table, it was found that the differences were in favor of males.

- No statistically significant differences at the significance level ($\alpha \leq 0.05$) were found on the fields of (scientific research and local community service) according to the educational qualification variable. Moreover, The (F) values were not statistically significant.
- Statistically significant teaching differences at the significance level ($\alpha \leq 0.05$) due to the educational qualification variable. The value of (F) reached (5.45), which is a statistically significant value. By referring to the arithmetic means shown in the previous table, it is clear that the differences are in favor of the academic qualification (Bachelor's degree).
- Statistically significant differences at ($\alpha \leq 0.05$) were found due to (teaching, scientific research, and local community service) according to a variable of teaching experience (more than ten years). Furthermore, the (F) values were statistically significant.
- The researchers explain this result to the fact that males have outperformed females in the abilities related to knowledge and research and standing actively in the search for information from its original sources, in addition to their ability to communicate with others to obtain information and adherence to confidentiality in their circulation and their investment in local events in raising awareness of the community members concerning the values of integrity and transparency. This reflects positively the affiliation and loyalty to the homeland and society by inviting the competent authorities to work on spreading the values and principles of academic integrity among members of society in general and school students in particular. The results of the study also indicated statistically significant differences for the educational qualification variable (bachelor's degree) for the field of teaching, and there were no statistically significant differences at the significance level ($\alpha \leq 0.05$) in the level of awareness of Arabic language teachers of the twelfth grade values of academic integrity according to the academic qualification variable in the two fields (scientific research and local community service). The researchers attribute this result to that teachers are

Table 6. The results the 3-way MANOVA to academic integrity values by (gender, educational qualification, and teaching experience)

Variable	Field	Total squares	Degree of freedom	Square mean	F	Statistical significance
Gender	Teaching	0.26	1	0.26	3.34	0.07
	Scientific research	0.63	1	0.63	7.91	0.01
	Community service	2.56	1	2.56	24.02	0.00
Academic qualification	Teaching	0.42	1	0.42	5.45	0.02
	Scientific research	0.29	1	0.29	3.64	0.06
	Community service	0.04	1	0.04	0.39	0.53
Teaching experience	Teaching	0.37	2	0.19	2.42	0.09
	Scientific research	0.26	2	0.13	1.61	0.20
	Community service	0.50	2	0.25	2.33	0.10
Error	Teaching	8.90	45	0.08		
	Scientific research	9.13	45	0.08		
	Community service	12.26	45	0.11		
Total	Teaching	2142.73	50			
	Scientific research	2111.35	50			
	Community service	2109.55	50			
Total (corrected)	Teaching	9.94	49			
	Scientific research	10.38	49			
	Community service	16.83	49			

very careful when teaching them to include what is related to the values of academic integrity by explaining the merits of the lesson or through the duties they assign to their students. Thus, they confirm their keenness to improve their academic roles in the institution to which they belong, just like their colleagues in higher academic qualifications.

7. Statistically significant differences at level ($\alpha \leq 0.05$) were evident in the level of awareness of Arabic language teachers of the twelfth grade, about the values of academic integrity due to the variable of experience. Such differences were in favor of the level of experience (more than 10 years). The researchers attribute this result to the fact that teachers who worked in the field of teaching for long periods dealt with the values of academic integrity more than others and employed them in a way that enhances the concept of motivation to accomplish tasks positively by their students. These findings are consistent with (Najm El-Din, 2017).

Third: Question Three: Is there a correlation between the level of awareness of Arabic language teachers of the twelfth grade concerning the values of academic integrity and motivation to complete tasks by their students?

The correlation coefficients were computed using the Pearson Correlation method between the mean of each teacher's ratings on the fields of academic integrity values and the arithmetic mean of the scores of his students in the achievement motivation measure. Table 7 illustrates this.

Table 7. Pearson Correlation coefficients between the arithmetic mean of each teacher's ratings on the fields of academic integrity values and the arithmetic mean of their students' scores in the achievement motivation questionnaire

Areas of academic integrity values	Questionnaire of motivation for achievement
Teaching	0.75*
Research	0.77*
Community service	0.74*

* Statistically significant at the significance level ($\alpha \leq 0.05$).

There is a clear correlation at the level ($\alpha \leq 0.05$) between the mean of each teacher's ratings on the fields of academic integrity values and the mean of the scores of his students in the motivational questionnaire to complete the tasks. The values of the correlation coefficients were statistically significant.

The researchers believe that teachers' awareness of their students about the values of academic integrity has been positively reflected on the motivation of their students to complete the tasks they have by assuming the responsibility to complete the work and duties required of them in addition to keeping pace with scientific and technological development in all fields of study and various school activities and excellence in competitions and teamwork, taking into account the principles of learning by trying and repetition in order to achieve mastery at work because of its positive impact in increasing the motivation for achievement among their students.

The two researchers attribute that teachers' awareness of the values of integrity and transparency in school work to the increase in the level of motivation to complete tasks by their students. It is reflected on their behavior and practices inside or outside the school and in their community and families. Thus, the student has a tendency towards achieving goals, which appears on his personality, and thus works to satisfy his internal

and external needs and desires to reach the desired goals with the help of a teacher who has sufficient awareness of the values of academic integrity. This result agreed with (Hamid and Kharbit 2018).

7 Recommendations and Suggestions

1. Creating a welcoming learning environment where students in the 12th grade can pursue their diverse interests, improve their skills, and invest their time in a way that upholds the principles of integrity and transparency.
2. Strengthening the relationship and extending bridges of cooperation and positive interaction between the school and other social institutions, to enhance the values of integrity and transparency.
3. Including the values of academic integrity in the curricula at all levels of education.
4. Conducting similar to the current study on students of public and private universities in the UAE.
5. Conducting studies addressing the necessity of introducing the values of integrity and transparency into educational policies and various curricula at all educational levels in public and private schools and universities, and their impact on society.

References

- Akram, H.A.M.S.: The effect of using the Telegram channel in enhancing the values of integrity among female forensic science teachers and their attitudes towards it. *Arabian Gulf message*. Arab Bureau of Education for the Gulf States (2017)
- Al Kurdi, B., Alshurideh, M., Al Afaishat, T.: Employee retention and organizational performance: evidence from banking industry. *Manag. Sci. Lett.* **10**(16), 3981–3990 (2020)
- Al Shraah, A., Abu-Rumman, A., Alqhaiwi, L., Alshurideh, M.T.: The role of AACSB accreditation in students' leadership motivation and students' citizenship motivation: business education perspective. *J. Appl. Res. High. Educ.* (ahead-of-print) (2022)
- Al-bawaia, E., Alshurideh, M., Obeidat, B., Masa'deh, R.: The impact of corporate culture and employee motivation on organization effectiveness in Jordanian banking sector. *Acad. Strat. Manag. J.* **21**(2), 1–18 (2022)
- Al-kharabsheh, S., Attiany, M., Alshawabkeh, R., Hamadneh, S., Alshurideh, M.: The impact of digital HRM on employee performance through employee motivation. *Int. J. Data Netw. Sci.* **7**(1), 275–282 (2023)
- Alsharari, N.M., Alshurideh, M.T.: Student retention in higher education: the role of creativity, emotional intelligence and learner autonomy. *Int. J. Educ. Manag.* **35**(1), 233–247 (2020)
- AlShehhi, Hind, Alshurideh, Muhammad, Kurdi, Barween Al, Salloum, Said A.: The impact of ethical leadership on employees performance: a systematic review. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *AISI 2020. AISC*, vol. 1261, pp. 417–426. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-58669-0_38
- Alshurideh, M.: Does electronic customer relationship management (E-CRM) affect service quality at private hospitals in Jordan? *Uncertain Supply Chain Manag.* **10**(2), 325–332 (2022)
- Alshurideh, M.T., Al Kurdi, B., Alzoubi, A.A., Akour, I.A.: The effectiveness of e-learning adoption on the education system during the Covid-19. *Int. J. Bus. Anal. Secur. (IJBAS)* **2**(1), 88–107 (2022)

- Alshurideh, M., Abuanzeh, A., Kurdi, B., Akour, I., AlHamad, A.: The effect of teaching methods on university students' intention to use online learning: technology acceptance Model (TAM) validation and testing. *Int. J. Data Netw. Sci.* **7**(1), 235–250 (2023)
- Alshurideh, M., Al Kurdi, B.H., Vij, A., Obiedat, Z., Naser, A.: Marketing ethics and relationship marketing—an empirical study that measure the effect of ethics practices application on maintaining relationships with customers. *Int. Bus. Res.* **9**(9), 78–90 (2016)
- Alshurideh, M., Al Kurdi, B., Abu Hussien, A., Alshaar, H.: Determining the main factors affecting consumers' acceptance of ethical advertising: a review of the Jordanian market. *J. Mark. Commun.* **23**(5), 513–532 (2017)
- Alshurideh, M., Kurdi, B.A., Shaltoni, A.M., Ghuff, S.S.: Determinants of pro-environmental behaviour in the context of emerging economies. *Int. J. Sustain. Soc.* **11**(4), 257–277 (2019)
- Alsuwaidi, Maryam, Alshurideh, Muhammad, Kurdi, Barween Al, Salloum, Said A.: Performance appraisal on employees' motivation: a comprehensive analysis. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020*, pp. 681–693. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-58669-0_61
- Alyammahi, Alyaa, Alshurideh, Muhammad, Kurdi, Barween Al, Salloum, Said A.: The impacts of communication ethics on workplace decision making and productivity. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *AISI 2020. AISC*, vol. 1261, pp. 488–500. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-58669-0_44
- Alzoubi, H., Alshurideh, M., Hussein, A., Kurdi, B., Ahmad, A., AlHamad, A.: Rationalizing the cost of quality through lean and agile operations practices: evidence from Aviation industry. *Uncertain Supply Chain Manag.* **11**(1), 319–324 (2023)
- Amarneh, B.M., Alshurideh, M.T., Al Kurdi, B.H., Obiedat, Z.: The impact of COVID-19 on E-learning: advantages and challenges. In: Hassanien, A.E., et al. (eds.) *AICV 2021. AISC*, vol. 1377, pp. 75–89. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_8
- Khan, B., Al-Anati, S., Abdulaziz, K., Al-Mobaideen, M.A.: The role of Jordanian universities in promoting integrity values and its relationship to national belonging among their students by applying to the University of Jordan and Al-Ahliyya Amman University. *Studies - Educational Sciences: University of Jordan - Deanship of Scientific Research* (2017)
- Balash, S., Kulthum.: Sources of occupational pressure and its relationship to motivation for achievement. *Al-Hikma J. Educ. Psychol. Stud.* **15**, 227 (2018)
- Bettayeb, H., Alshurideh, M.T., Al Kurdi, B.: The effectiveness of Mobile learning in UAE universities: a systematic review of motivation, self-efficacy, usability and usefulness. *Int. J. Control Autom.* **13**(2), 1558–1579 (2020)
- The American University of Beirut. *Academic Integrity: A matter that matters to us. Academic Principles: Accredited by the American University in Cairo Board* (2003)
- Al-Jwhary, A.N., bin Hammad, I.: *The Right Crown and Arabic Language*, 4th edn. Dar Al-Ilm for Millions, Beirut (1987)
- Al-Harbi, M.M.A.: The degree of commitment to the practice of administrative transparency in the academic departments in the College of Education at King Saud University. *Spec. Int. Educ. J.: Dar Simat Stud. Res.* **1**(6), 308 (2012)
- Al-Harbi, M.A.: Determinants of violating academic integrity standards among undergraduate and postgraduate students in the Kingdom of Saudi Arabia. *J. Human. Soc. Sci.* **39**, 209 (2016)
- Hussein, A.R., Al-Safi, A.H., et al.: *Motivation and emotional intelligence*, Amman: Dar Al-Fikr for Publishing and Distribution, 1st edn (2006)
- Dweik, A.G.A.: *Modern methods used in educational institutions in protecting integrity and combating corruption*. Naif Arab University for Security Sciences (2013)
- Khalifa, A.A.O.: A program to develop children's achievement motivation. *Arab J. Social Sci.* **4**(3), 12 (2013)

- Khuwailid, A.: Measurement of achievement motivation. *Al-Hikma J. Educ. Psychol. Stud.* **2016**(8), 298 (2016)
- Khuwailid, A.: Theoretical context for the emergence of the concept of motivation for achievement Studies: Ammar Tleji University, Laghouat (2016)
- Kurdi, B., Alquqa, E., Alzoubi, H., Alshurideh, M., Al-Hawary, S.: The effect of process quality improvement and lean practices on competitive performance in the UAE healthcare industry. *Uncertain Supply Chain Manag.* **11**(1), 261–266 (2023)
- Kurdi, B., Alshurideh, M., Alnaser, A.: The impact of employee satisfaction on customer satisfaction: theoretical and empirical underpinning. *Manag. Sci. Lett.* **10**(15), 3561–3570 (2020)
- Leo, S., Alsharari, N.M., Abbas, J., Alshurideh, M.T.: From offline to online learning: a qualitative study of challenges and opportunities as a response to the COVID-19 pandemic in the UAE higher education context. In: Alshurideh, M.T., Hassanien, A.E. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence*. SSDC, vol. 334, pp. 203–217. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67151-8_12
- Nuseir, Mohammed T., Al, Barween H., Kurdi, Muhammad T., Alzoubi, Haitham M.: Gender Discrimination at Workplace: Do Artificial Intelligence (AI) and Machine Learning (ML) Have Opinions About It. In: Hassanien, Aboul Ella, Haqiq, Abdelkrim, Tonellato, Peter J., Bellatreche, Ladjel, Goundar, Sam, Azar, Ahmad Taher, Sabir, Essaid, Bouzidi, Driss (eds.) *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2021)*, pp. 301–316. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_28
- Shukri, A.Y.: *Specialized International and Regional Organizations*, p. 282. Iturk for Publishing and Distribution, Cairo (2003)
- Al-Shanqiti, M.H.M.: The educational role of the university to consolidate integrity in society from the point of view of the faculty members at the University of Tabuk. *Educ. J.* **162**(1), 271 (2015)
- Abdel-Saheb, M.M., Ahmed, M.F., Faisal, M.: Academic integrity in scientific research among faculty members at the University of Baghdad from their point of view. *J. Islamic Res. Stud.* **54**, 309 (2016)
- Afana, G.B.I., Al-Waeli, S.A.K.: The effect of teaching according to the two types of audio-visual learning in developing oral expressive performance and motivation for achievement in the Arabic language among students of the basic stage in Zarqa: the fourth grade as a model (unpublished master's thesis). The Hashemite University, Zarqa (2013)
- Farah, S., et al.: *Curriculum Development in the United Arab Emirates*, Sheikh Saud bin Saqr Al Qasimi Foundation for Public Policy Research, Public Policy Paper No. 18 (2017)
- Fadloun, Z.: Total quality management as one of the mechanisms for promoting a culture of academic integrity in the university institution. *J. Hum. Sci.* **8**, 111 (2017)
- Qatami, Y., Adass, A.R.: *General Psychology*, Dar Al-Fikr for printing and publishing (2017)
- Muhammad, M.F.A.: Attitudes of faculty members in Egyptian universities towards plagiarism: a field study. *Int. J. Libr. Inf. Sci.* **6**(2), 13 (2019)
- Al-Mazrouei, S., et al.: *The General Framework for Education Policies in the United Arab Emirates*, 1st edn. Ministry of Education (2018)
- Mansi, H.: *Classroom Administration - Jordan* Dar Al Kindi for Publishing and Distribution, 2nd edn (2000)
- Melhem, S.: *The psychology of learning and teaching Theoretical and applied foundations*. Dar Al Uloom for investigation, printing, publishing and distribution (2006)
- The Emirati Education System, United Arab Emirates: Ministry of Education, 2015–2017
- Al-Din, N., Hanan, A.-J.A.-G.: The effect of using classroom activities to enhance the values of integrity in teaching among female student teachers at the General Diploma in Education at the University of Jeddah. *Arab Stud. Educ. Psychol.* (2017)

- Rahmi, Y., Bilal. S.: The effects of accountant professionals' social responsibility perceptions on self esteem and job performance. *Ataturk Universities Iktisadi Vet Dari Bilimler Dergisi, Cilt: 26*, 3–4 (2012)
- Spaulding, C.L.: *Motivation in the Classroom*. McGraw-Hill, New York (1992)

Data Sciences and Business Based Applications



Short-Term Forecasting of GDP Growth for the Petroleum Exporting Countries Based on ARIMA Model

Sara Abdelghafar¹(✉), Ashraf Darwish², and Abdulrahman Ali³

¹ School of Computer Science, Canadian International College (CIC), Cairo, Egypt
sara_abdelghafar@cic-cairo.com

² Faculty of Science, Helwan University, Helwan, Egypt

³ RiskSpan, Arlington, USA

Abstract. Petroleum is an essential part of all aspects of today's technology, and its consumption is strongly correlated with economic growth. This paper presents a forecasting model of petroleum GDP (Gross Domestic Product) for the members of OPEC (Organization of the Petroleum Exporting Countries). The proposed model is presented with the statistical ARIMA (Autoregressive Integrated Moving Average) model which is considered one of the most effective methods for forecasting stationary time-series data. The proposed model is mainly based on the penalized likelihood method used to obtain the optimal ARIMA parameters. The obtained results show that the proposed model is capable of achieving high accuracy for prediction and short-term forecasting for this type of time series data. Therefore, the proposed model has the potential to be used as an important tool in forecasting petroleum OPEC Members' GDP, which will facilitate proper monitoring and control of OPEC petroleum consumption and economic growth.

Keywords: ARIMA · GDP · OPEC · Petroleum economic growth · Time-series · Forecasting · Penalized likelihood

1 Introduction

Without energy, human activities in today's world would be impossible. Energy is the lifeblood of modern technology and its operations, and it is closely tied to economic development [1]. Petroleum and gas provide around two-thirds of the world's energy. Petroleum consumption is positively linked to economic growth, implying that the more developed an economy is, the more petroleum is consumed [2].

Over the past decades, the Organization of the Petroleum Exporting Countries (OPEC) is being played a vital role in sustaining global economic activity. The importance of OPEC member countries is shown in supplying the world with essential energy by producing and exporting petroleum [3]. OPEC is an intergovernmental organization, it was founded in Baghdad in September 1960 with five founding members; Iraq, IR Iran, Saudi Arabia, Kuwait, and Venezuela. Currently, the OPEC headquarters is located

in Vienna, Austria, it comprises 13 member countries; Algeria, Angola, Congo, Equatorial Guinea, Gabon, IR Iran, Iraq, Kuwait, Libya, Nigeria, Saudi Arabia, United Arab Emirates, and Venezuela [4].

Autoregressive Integrated Moving Average (ARIMA) is one of the most effective forecasting methods that have been widely used in many domains [5, 6]. In this study, the proposed ARIMA model is used to forecast the total growth economy domestic for OPEC member countries. Originally, ARIMA models were based mainly on using the autocorrelation function (ACF) and partial autocorrelation function (PACF) to identify the characteristics and stationarity of the time-series data and then identify the model parameters. However, manually finding the optimal parameters for ARIMA using ACF and PACF plots can be time-consuming because the number of models to evaluate is a permutation of the number of model order parameters. Therefore, the proposed model uses Akaike information criteria (AIC) and Bayesian information criteria (BIC) as the penalized-likelihood criteria to obtain the best model parameters by finding the lowest AIC and BIC values. The prediction evaluation metric is used to evaluate the prediction accuracy of the best models obtained, as will be analyzed and discussed later in the following sections. The obtained results prove the capability of the proposed model for achieving high accuracy for this type of time series data.

The rest of the paper is organized as follows: Sect. 2 describes the ARIMA model, both seasonal and non-seasonal models. The proposed framework of the short-term forecasting model parameter selection and optimization procedure is discussed in Sect. 3. In Sect. 4, the obtained results are discussed and analyzed. Finally, Sect. 5 presents the conclusions and future work.

2 Auto-regressive Integrated Moving-Average

The autoregressive integrated moving average model is a generalized form of an autoregressive moving average (ARMA) model. ARIMA is one of the most effective methods for forecasting stationary time-series data [7, 8]. ARIMA model is characterized by three main parameters (p, d, q) which can be defined as follows, (p) is the order of the Auto Regression (AR) term that is the regression of a specific variable against itself to forecast the variable of interest. Whereas the (q) is the order of the Moving Average (MA) term that uses the errors associated with the forecast at a previous time step to forecast a variable at a later time step. The “I” symbol in ARIMA refers to the integration of AR and MA models using the differencing term (d) [6]. Equation 1 and Eq. 2 are the generalized equations of p^{th} order of AR model and q^{th} order of MA model, respectively:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

$$y_t = C + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2)$$

So the integration process of the ARIMA model can be mathematically represented as in Eq. 3.

$$y_t = C + \phi_1 y + \phi_p y_{t-p} + \dots + \phi_n y_{t-n} + \theta_1 \varepsilon_{t-1} + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

where n is number of observations, C is constant, $\varphi_i (i = 1, 2 \dots p)$ is AR model parameters, $\theta_i (i = 1, 2 \dots p)$ is MA model parameters, y_t is current time-series value, $y_{t-1}, y_{t-2} \dots y_{t-p}$ is past values and ε_t is a random error or residual term for the t^{th} time.

The seasonal part of an ARIMA model is Seasonal-ARIMA (SARIMA), which includes non-seasonal ARIMA (p, d, q) and additional seasonal parameters (P, D, Q)_m which are the order of seasonal AR term, seasonal MA term, and seasonal differencing term, for time-series data for m number of time steps corresponding to a single seasonal period [9]. The mathematical model of SARIMA is represented by Eq. 4 as follows:

$$\varphi_P(B^m)\varphi_P(B)(1-B^m)^D(1-B)^d y_t = \vartheta_Q(B^m)\theta_q(B)\omega_t \quad (4)$$

where y_t is the time series, ω_t is the Gaussian white noise process, $\varphi_P(B)$ is non-seasonal auto-regressive polynomial and $\theta_q(B)$ is non-seasonal moving average polynomial, d is the non-seasonal differencing term where D is the seasonal differencing term, $\varphi_P(B^m)$ is seasonal auto-regressive polynomial, $\vartheta_Q(B^m)$ is seasonal moving average polynomial, and B is defined as the backshift operator [10].

3 Proposed Short-Term Forecasting Model

The proposed short-term forecasting model is initialized with the ACF and PACF, which are used to indicate the relation between the observations within the time series to identify their characteristics and stationarity and to estimate the initial parameters of the forecasting model. Where the ACF is a natural generalization of the linear correlation coefficient that is used to measure the correlation of the time series with its previous time-series data to identify the moving-average model. While PACF correlates the time series with its own lagged values separated by certain time units for identifying the autoregressive model. For the time series y_t the ACF and PACF are mathematically represented by Eq. 5 and Eq. 6 respectively [11].

$$ACF(y_t, y_{t-k}) = \frac{Covariance(y_t, y_{t-k})}{variance(y_t)} \quad (5)$$

$$PACF(y_t, y_{t-k}) = \frac{Covariance(y_t, y_{t-k} | y_{t-k+1})}{\sqrt{variance(y_t | y_{t-k+1})} \sqrt{variance(y_{t-k} | y_{t-k+1})}} \quad (6)$$

where k is lag, and it is defined as the difference between y_t and y_{t-k} .

After identifying the potential models using the ACF and PACF functions, the optimization and selection processes based on the AIC and BIC with the penalized likelihood criteria are used to obtain the optimal model for the used time-series data.

$$AIC = -2\log L(\hat{\theta}) + 2N \quad (7)$$

$$BIC = -2\log L(\hat{\theta}) + N\log n \quad (8)$$

where $\log L(\hat{\theta})$ represents the likelihood function, N is the total number of model parameters and n is the number of the observations.

The evaluation metrics used in this approach are mean square error (MSE), root mean square error (RMSE), and mean absolute percentage error (MAPE) which are used often to evaluate the prediction accuracy [5, 12], as given by Eq. 9, Eq. 10, and Eq. 11. The flowchart of the proposed ARIMA short-term forecasting model is depicted in Fig. 1, showing the procedures of selection of the model parameters and evaluation.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{10}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{11}$$

where $\hat{y}_i$ is the model predicted value, y_i is the actual value.

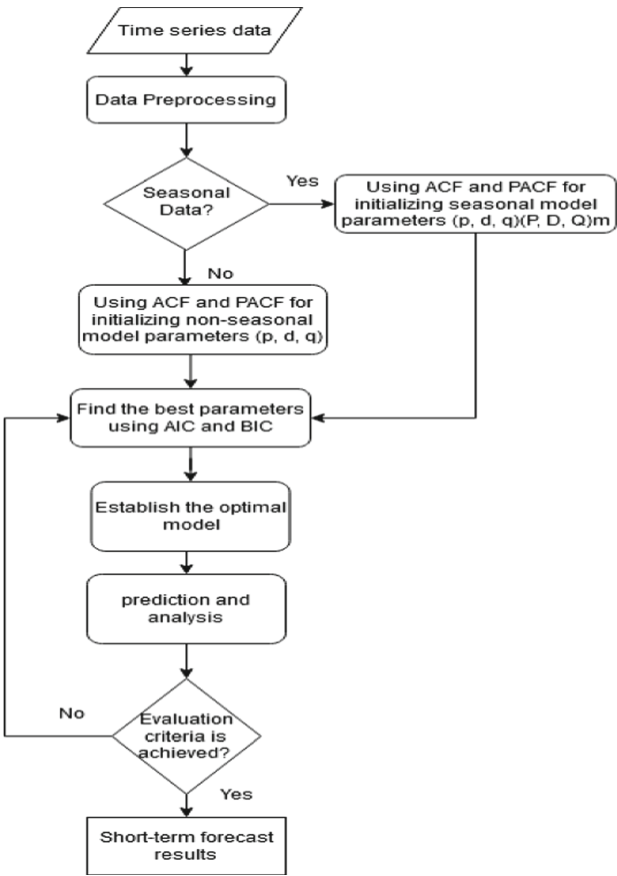


Fig. 1. The proposed ARIMA short-term forecasting model

4 Results and Discussion

The OPEC Members' GDP dataset is obtained from the OPEC ASB (annual statistical bulletin), which is freely available online on the OPECs official website [13]. The dataset is collected from the total OPEC Members' GDP in millions of dollars yearly from 1960 to 2020, for the thirteen countries, as depicted in Fig. 2.

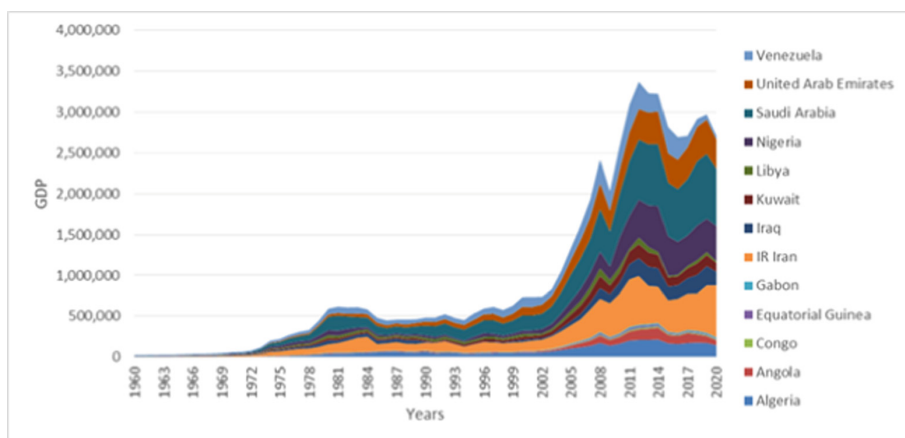


Fig. 2. OPEC members' GDP at current market prices (m \$)

The dataset is normalized to improve the accuracy and convergence speed. Then the correlation coefficient has been used to investigate the relationship between the independent variables and the target forecasting value. Table 1 shows all correlations with positive values, which indicate that the domestic petroleum economic growth of OPEC member countries and the total OPEC petroleum domestic economic growth have moved in the same direction. Therefore, the petroleum domestic economic growth of each of the 13 OPEC member countries is the input, and the total growth economy domestic for OPEC is the target forecasting of the proposed ARIMA model.

The time-series data of the 13 countries were split into 80% training and 20% testing/validating datasets. The model development and parameter optimization were done using the training dataset and the performance of the developed model was tested with the validation dataset. A confidence level of 95% is established for the prediction intervals. The best model is selected based on the prediction error criteria MSE, RMSE, and MAPE.

Stationarity is an essential condition for modeling and forecasting, so ACF and PACF plots have been used to check the stationarity of the time-series data. The results for the raw time-series data show an autocorrelation reduces very marginally as the number of lags increases, which indicates that the time series is non-stationary, as depicted in Fig. 3 (a) and (b). As the presence of non-stationary time series data has been confirmed, the differencing must be performed before the ARIMA or SARIMA modeling. So the non-stationary raw data were subjected to first-order differencing (I or $d = 1$) to stabilize the mean of the time series before performing the forecast. Also, no need for a higher

Table 1. Correlation relationship between OPEC petroleum domestic economic growth and individual member countries.

Country	OPEC	Country	OPEC
Algeria	0.985	Kuwait	0.984
Angola	0.970	Libya	0.672
Congo	0.985	Nigeria	0.967
Gabon	0.982	Saudi Arabia	0.993
IR Iran	0.976	Venezuela	0.840
Iraq	0.965	United Arab Emirates	0.989
Equatorial Guinea	0.946		

order of differencing, as there is no seasonality and no positive autocorrelations out to a high number of lags, as depicted in Fig. 3 (c) and (d).

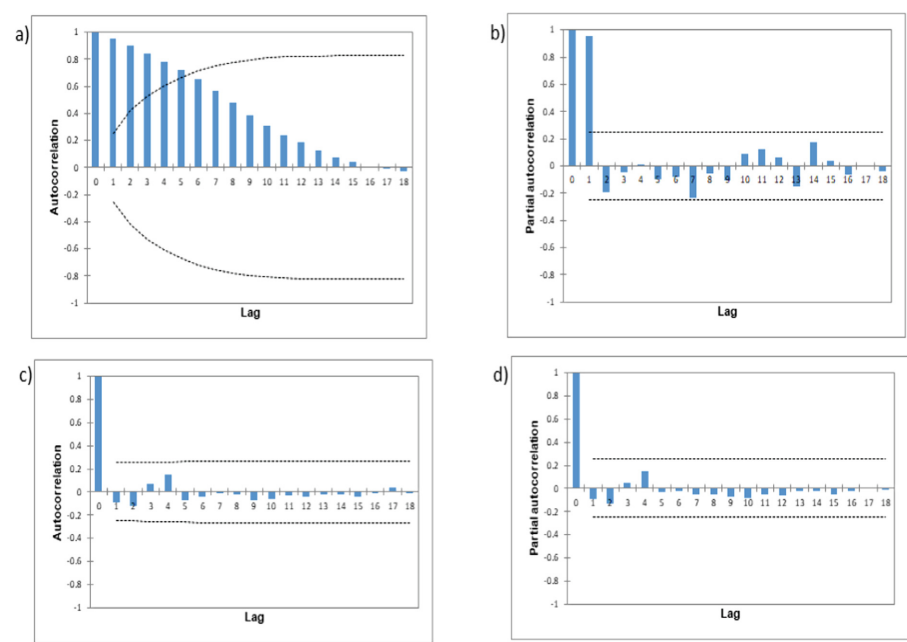


Fig. 3. (a) ACF plot of the raw time series data, (b) PACF plot of the raw time series data, (c) ACF plot of the time series data after first differencing, (d) PACF plot of the time series data after first differencing.

After the ARIMA model parameters have been initialized with ACF and PACF, the best model was identified using the penalized likelihood criteria Table 2. Shows the obtained best models that have the smallest prediction errors based on the lowest AIC

and BIC values. Once the best model is identified by training on the training dataset, the model was used to predict values of the test data followed by forecasting for the next 10 years of the total OPEC members' GDP. R-Squared (R^2) is used to show the goodness of fit for the forecasting results obtained from the best prediction models. As depicted in Fig. 4, the actual values are plotted against the predicted values of the best model (2, 1, 5), then the forecasted results for the next 10 years using this model.

Table 2. The best ARIMA models and prediction error criterion (MSE, RMSE, MAPE, and R^2)

Model	MSE	RMSE	R2	MAPE
ARIMA (2,1,5)	7.65E-08	0.000277	0.999999993	0.0376
ARIMA (1,1,5)	7.96345E-08	0.000282	0.999999992	0.0471
ARIMA (1,1,1)	8.24489E-08	0.000287	0.999999992	0.0341
ARIMA (2,1,2)	9.21082E-08	0.000303	0.999999991	0.0541
ARIMA (4,1,2)	1.92819E-07	0.000439	0.999999982	0.1364
ARIMA (5,1,7)	2.29472E-07	0.000479	0.999999979	0.1966

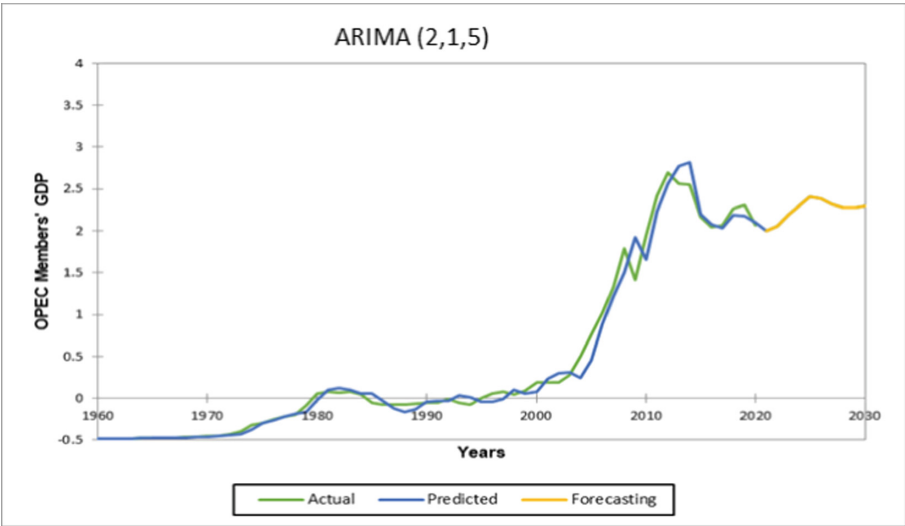


Fig. 4. The best ARIMA model (2, 1, 5) for OPEC members' GDP prediction and forecasting

5 Conclusion and Future Work

In this paper, we have forecasted the economic growth for the petroleum exporting countries for ten years using the ARIMA statistical model. Initial ARIMA model parameters

were selected using the analytical tools ACF and PACF and then followed by finding the optimized model parameters based on the lowest value of AIC and BIC that were used to assess the reliability of the models. Predication evaluation metrics MSE, RMSE, and MAPE were used as criteria for selecting the best model. As well, R² was used to show the goodness of fit for the forecasting results obtained from the best prediction models. The obtained results show that the proposed model is capable to achieve high accuracy for prediction and short-term forecasting for this type of time series data. The future work of this paper will focus on applying the proposed model to other types of series time data including the seasonality conditions. Moreover, we plan to improve the performance of the proposed model using evolutionary or swarm-based optimization algorithms.

References

1. Suganthi, L., Samuel, A.A.: Energy models for demand forecasting—a review. *Renew. Sustain. Energy Rev.* **16**(2), 1223–1240 (2012)
2. Chiroma, H., et al.: A new approach for forecasting OPEC petroleum consumption based on neural network train by using flower pollination algorithm. *Appl. Soft Comput.* **48**, 50–58 (2016)
3. Khan, A., et al.: Forecasting electricity consumption based on machine learning to improve performance: a case study for the organization of petroleum exporting countries (OPEC). *Comput. Electr. Eng.* **86** (2020). <https://doi.org/10.1016/j.compeleceng.2020.106737>
4. OPEC. <https://www.opec.org>. Accessed 11 Feb 2022
5. Tylkowski, J., Hoja, M.: Time decomposition and short-term forecasting of hydrometeorological conditions in the south Baltic coastal zone of Poland. *MDPI Geosci. J.* **68**(2), 1–18 (2019). <https://doi.org/10.3390/geosciences9020068>
6. ArunKumar, K.E., Kalaga, D.V., Kumar, C.M.S., Chilkoor, G., Kawaji, M., Brenza, T.M.: Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: auto-regressive integrated moving average (ARIMA) and seasonal auto-regressive integrated moving average (SARIMA). *Appl. Soft Comput. J.* **103**, 107–161 (2021). <https://doi.org/10.1016/j.asoc.2021.107161>
7. ArunKumar, K.E., Kalaga, D.V., Kumar, C.M.S., Kawaji, M., Brenza, T.M.: Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alexandria Eng. J.* **61**, 7585–7603 (2022). <https://doi.org/10.1016/j.aej.2022.01.011>
8. Jai, G., Mallick, B.: A study of time series models ARIMA and ETS. *J. Mod. Educ. Comput. Sci* **4**, 57–63 (2017)
9. Chen, P., Niu, A., Liu, D., Jiang, W., Ma, B.: Time series forecasting of temperatures using SARIMA: an example from Nanjing. In: *IOP Conference Series: Materials Science and Engineering*, Chen, Peng (2018). <https://doi.org/10.1088/1757-899x/394/5/052024>
10. He, K., Ji, L., Wu, C.W.D., Tso, K.F.G.: Using SARIMA–CNN–LSTM approach to forecast daily tourism demand, vol. 49, pp. 25–33 (2021). <https://doi.org/10.1016/j.jhtm.2021.08.022>
11. Zohair, M., et al.: ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. *Neural Comput. Appl.* **33**(7), 2929–2948 (2021)
12. Abdelghafar, S., Goda, E., Darwish, A., Hassanien, A.E.: Satellite lithium-ion battery remaining useful life estimation by coyote optimization algorithm. In: *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 124–129 (2019). <https://doi.org/10.1109/ICICIS46948.2019.9014752>
13. OPEC Annual Statistical Bulletin (ASB) (2021). <https://asb.opec.org/>. Accessed 10 Jan 2022



Technique for Order of Preference by Similarity to Ideal Solution Based Methodology for Detecting Important Actors in Social Networks: Facebook Ego Network as a Case Study

Khaoula Ait Rai^{1(✉)}, Mustapha Machkour¹, and Jilali Antari²

¹ Computer System and Vision Laboratory, Faculty of Sciences, Ibn Zohr University,
BP 8106, Agadir, Morocco

khaoula.aitrai@edu.uiz.ac.ma

² Laboratory of Computer Systems Engineering, Mathematics and Applications,
Polydisciplinary Faculty of Taroudant, Ibn Zohr University, BP 8106, Agadir, Morocco

Abstract. Detecting important actors in social networks is still a hot research topic in the complex network area. Various metrics of centrality are proposed to deal with this problem. Each metric addresses the problem from one point of view. Therefore, this paper introduces a new method to estimate the influence of nodes using multi-criteria decision-making MCDM. The proposed approach is based on the technique for order preference by similarity to an ideal solution (TOPSIS). TOPSIS aggregates centrality measures. Degree, betweenness, closeness, and eigenvector centralities are combined in this work to detect the importance of each actor. Each metric of centrality has its limitations, thus the proposed approach takes advantage of these four centralities together through TOPSIS. To illustrate the effectiveness of the proposed methodology, the Facebook ego network as an example of a social network is used for experimentation. The SI model is used to compare the spreading ability of nodes between centralities and the proposed approach on the Facebook ego network. The experimental results show that the proposed methodology can rank the important nodes more efficiently than a single measure of centrality.

Keywords: Complex networks · TOPSIS · Important actors · Centrality measures · SI model

1 Introduction

Complex networks are ubiquitous in all areas of life such as computer sciences, biological sciences, and social sciences (Mendes et al. 2012; Bullmore and Sporns 2009). Networks consist of a set of nodes or actors in the case of social networks. Nodes in complex networks can be computers, proteins, web pages, cities, etc. These nodes don't

have the same role in the network. Some nodes can affect the whole structure by diffusing information throw the network compared to other remaining nodes. These nodes are called “important nodes”. How identify important nodes and evaluating the influence of each node is a very challenging topic in network science and has various applications. For instance, it can be used in the prevention of the case of transmission of epidemics (Barabási et al. 2011), stop spreading rumors (Xiong and Hu 2012), and keep the connection in systems (Resende and Pardalos 2006). This issue of node’s importance got a lot of attention from researchers in the last decade. Many centrality measures are proposed in the past to solve this issue such as degree centrality (Bródka et al. 2011), betweenness centrality (Brandes 2001), closeness centrality (Du et al. 2015), Katz centrality (Katz 1953), PageRank (Cheng et al. 2009), etc. Each measure has a point of view about the node’s importance. Thus, there is no consensus on the best one. Degree centrality for example considers how many links each node is holding. It’s the simplest measure of node connectivity but doesn’t consider the whole structure of networks. The amount of times a node is on the shortest path between other nodes is measured by betweenness centrality. Closeness centrality relies on the node’s closeness to all other nodes in the network to influence the entire network quickly. In this metric, a score for each node is assigned based on the sum of the shortest path. The complexity in closeness and betweenness centralities is the calculation of all the shortest paths in the network which makes the task delicate in large-scale networks. EigenVector centrality (Bonacich and Lloyd 2001) is a variant of degree centrality. It goes a step even more by additionally evaluating a node’s degree of connectivity by considering the extended connections of a node.

As long as there is no consensus to choose the best measure, these four measures mentioned above are assembled via a multi-criteria analysis method (Chen and Hwang 1992) called TOPSIS. Multi-criteria analysis is used to make a comparative judgment between heterogeneous measures (Chen and Hwang 1992). It’s applied to many areas. TOPSIS (Galam and Zucker 2000) is a node ranking method. That chooses a certain number of alternatives based on a set of favorable or unfavorable criteria. This method is part of the techniques used in the field of multi-criteria decision support. It was developed by Hwang and Yoon in 1981 (Chen and Hwang 1992). Its principle consists in determining for each alternative a coefficient between 0 and 1 based on the distances (Euclidean) between each alternative on the one hand and the favorable and unfavorable ideal solutions. Section 2 below explains the steps to follow one by one.

In this paper, the detection of influential nodes is treated by applying TOPSIS. Degree, betweenness, closeness, and eigenvector centrality are used as a multi-attribute in TOPSIS. Firstly, centrality measures are calculated to be used as an input of TOPSIS. Then, the TOPSIS process generates a ranking list of nodes and finally, the most important nodes in the network are detected.

The paper is organized as follows. Section 2 presents the proposed methodology. The experimental results are highlighted in Sect. 3. Finally, some conclusions are drawn in Sect. 4.

2 Proposed Methodology

The methodology introduced in this paper is based on TOPSIS (Chen and Hwang 1992). This methodology aggregate some centrality measures as a multi-attribute. These measures are degree centrality DC (Bródka et al. 2011), betweenness centrality BC (Brandes 2001), closeness centrality CC (Brandes et al. 2016), and eigenvector centrality EC (Bonacich and Lloyd 2001). Each measure has its pros and cons as mentioned in the previous section. This methodology based on TOPSIS has allowed us to take advantage of each metric. It's a multi-criteria decision-making method MCDM.

The proposed methodology as explained in Fig. 1 is applied to a social network. Node's centrality (DC, BC, CC, EC) is calculated and used to form the TOPSIS decision matrix. In the TOPSIS process as explained in Fig. 2, normalization and weighting of the matrix are necessary. Then, the calculation of the positive ideal solution and the negative ideal solution. The calculation of the Euclidean distance from the positive and negative ideal solution, the calculation of the performance score which determines the relative proximity to the ideal solution, and finally the ranking of the nodes.

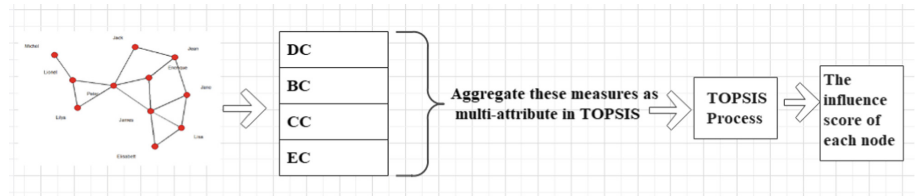


Fig. 1. The flowchart of the proposed approach.

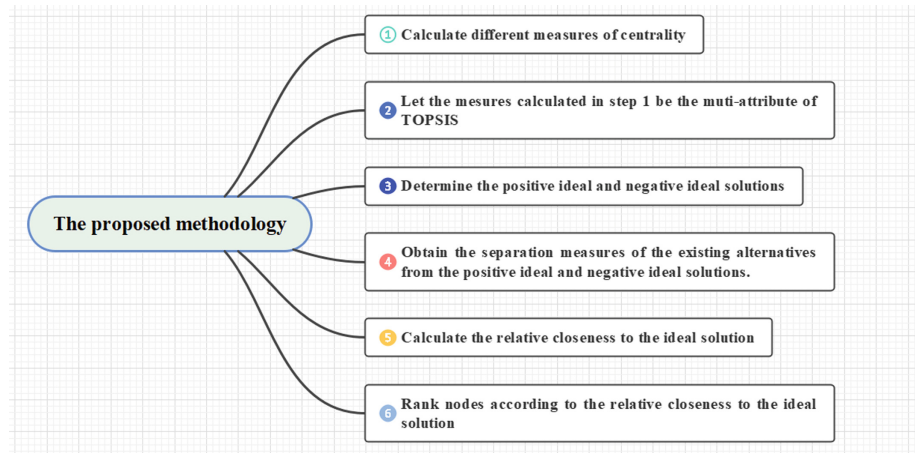


Fig. 2. TOPSIS process.

The steps of the proposed methodology are detailed in Algorithm 1 below:

Algorithm1. Steps of the proposed methodology.

Input: A social network (V, E) , m measures of centrality.

Output: top-ten influential nodes.

1. Create a decision matrix P based on m measures of centrality where $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$ and n is the number of nodes.

2. Normalized the decision matrix P

$$P = \begin{bmatrix} d_{11} & \cdots & d_{1m} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{nm} \end{bmatrix}$$

3. Make the decision matrix normalized and weighted to obtain the matrix W

$$W = \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{nm} \end{bmatrix}$$

4. Calculate the ideal best solution and negative ideal solution (centrality measures are taken as benefit attributes)

$$B^+ = \max (w_{ij}) \text{ and } B^- = \min (w_{ij})$$

5. Calculate separation between ideal and negative Solutions

$$Sep_p = \sqrt{\sum_{j=1}^m (w_{ij} - B^+)^2} \text{ and } Sep_n = \sqrt{\sum_{j=1}^m (w_{ij} - B^-)^2}$$

6. Calculate the relative closeness score of each node

$$\text{For each } w_{ij} \text{ in } W \text{ do } X_i = \frac{Sep_n}{Sep_n + Sep_p}$$

7. Rank nodes based on X_i the higher score means important node: top-ten influential nodes

3 Experimental Results

3.1 Dataset

To evaluate the performance of the suggested method in this paper, a Facebook network dataset is used to show its efficacy. This dataset consists of ‘circles’ (or ‘friends lists’) from Facebook (Leskovec and Mcauley 2012). Facebook data was collected from survey participants using a Facebook app. The dataset includes node features (profiles), circles, and ego networks. It is available at <https://snap.stanford.edu/data/ego-Facebook.html>. Table 1 below presents the characteristics of this dataset and Fig. 2 presents the graph. Nodes are presented with the blue color and links between them with black. This work is implemented under Python using these libraries: Math, Numpy, Pandas, NetworkX, Matplotlib, and NDlib for simulating the diffusion process.

Table 1. Characteristics of the Facebook dataset (“SNAP: Network datasets: Social circles,” n.d.)

Nodes	4039
Edges	88234
Nodes in largest WCC	4039 (1.000)
Edges in largest WCC	88234 (1.000)
Nodes in largest SCC	4039 (1.000)
Edges in largest SCC	88234 (1.000)
Average clustering coefficient	0.6055
Number of triangles	1612010
Fraction of closed triangles	0.2647
Diameter (longest shortest path)	8
90-percentile effective diameter	4.7

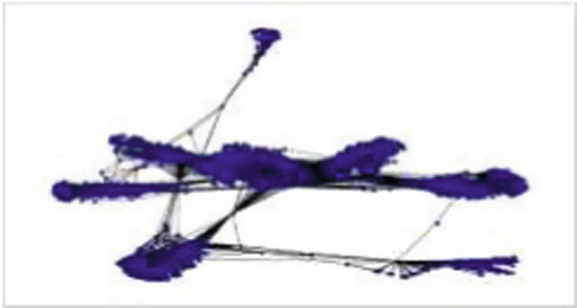


Fig. 3. The visualization of the dataset Facebook ego.

3.2 Experiments and Effectiveness

The implementation of the proposed methodology consists of constructing the network, computing and normalizing centrality measures, and computing the relative closeness score. The top ten influential nodes depending on the relative closeness are highlighted in Table 2 from the largest value to the smallest value.

To evaluate the efficiency of the TOPSIS method, the influence of the top 10 influential nodes obtained by the TOPSIS method is compared with other centrality measures using the SI model. The following table shows the top 10 influential nodes by different centrality measures and the TOPSIS-based approach.

Table 2. Top 10 influential nodes detected using TOPSIS.

Node	Score
107	0,913277
1684	0,695566
1912	0,496063
3437	0,488865
0	0,304321
1085	0,297379
698	0,231106
567	0,193646
58	0,169464
428	0,132923

Table 3. The top 10 influential nodes by different centrality measures and the proposed approach.

Ranking	DC	BC	CC	EC	Proposed approach
1	107	107	107	1912	107
2	1684	1684	58	2266	1684
3	1912	3437	428	2206	1912
4	3437	1912	563	2233	3437
5	0	1085	1684	2464	0
6	2543	0	171	2142	1085
7	2347	698	348	2218	698
8	1888	567	483	2078	567
9	1800	58	414	2123	58
10	1663	428	376	1993	428

According to Table 3, the TOPSIS-based approach and DC have the same five members in the top 10 lists, the TOPSIS-based approach and BC have the same top-ten lists, the TOPSIS-based approach and CC have the same three members in the top-ten lists, the TOPSIS-based approach and EC have the same one member in the top 10 lists.

The SI model will be used for comparison between the TOPSIS-based approach and centrality metrics. SI model is the simple deterministic model. This is the model that W.H. Hamer developed in 1906 (Hamer 1906). It initially postulates that there is no death and no recovery. It has two states: Susceptible (S) or infected (I). In addition, the populations of healthy individuals and infected individuals are constantly in contact. The infection is established by direct contact between an infected individual and a healthy individual. A flow is established between S and I. It depends on the number of healthy

individuals, the number of contagious individuals, and a proportionality coefficient β called the infection rate or contagion rate. The existence of this rate means that all contacts between healthy and contagious do not necessarily take place during the time interval and that a contact is not systematically contaminating. Therefore, β depends on many biological but also sociological or economic factors.

The spreading ability of nodes that either appear in the top 10 list by the proposed method or DC, BC, CC, and EC are compared in this section. The implementation of the cumulative infected nodes for the Facebook network is shown in Fig. 4 and Fig. 5. In this paper, two different values of (Infection probability) are used: 0.3 and 1. In each iteration, a set of nodes will be infected randomly by a probability of 0.3 and 1 respectively. To increase the precision of the results, the algorithm will be repeated 10000 times. Thus, the variation (standard deviation) and the mean of each iteration will be calculated. Figure 4 and Fig. 5 present the results.

According to Fig. 4 and Fig. 5, it can be clear that the TOPSIS method performs better than DC, CC, and EC, and as well as BC which is normal because the top 10 nodes of the TOPSIS-based approach and BC are the same.

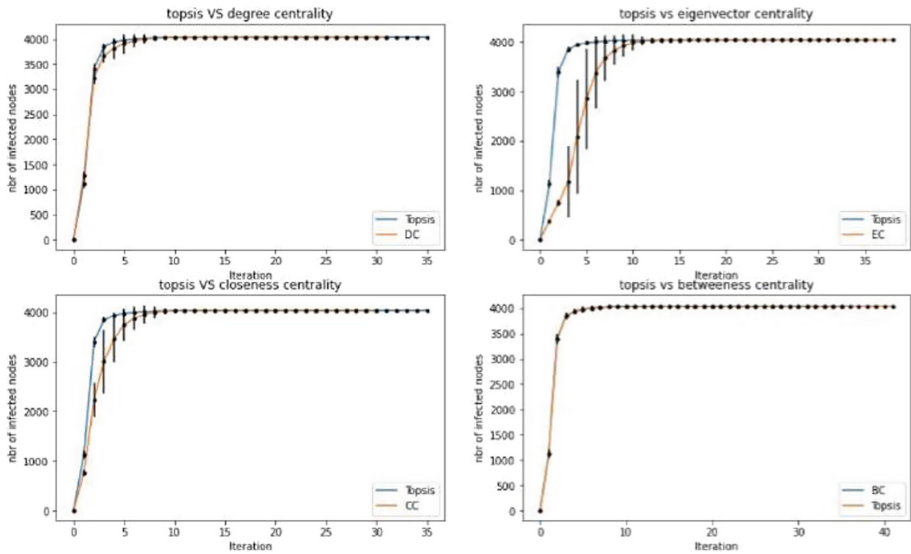


Fig. 4. The cumulative number of infected nodes as a function of time in the Facebook dataset with $\beta = 0.3$.

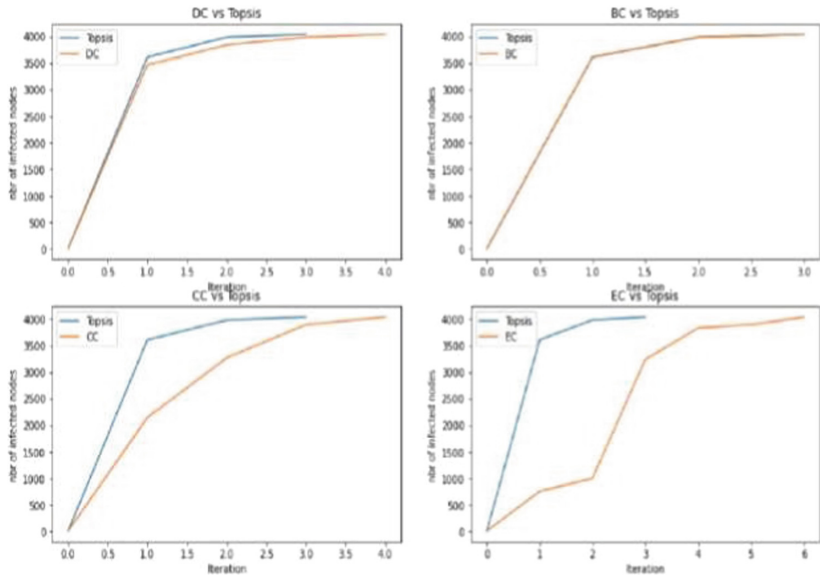


Fig. 5. The cumulative number of infected nodes as a function of time in the Facebook dataset with $\beta = 1$.

4 Conclusion

Complex networks are ubiquitous in our real life. Detecting the importance of nodes is still an open issue in network science. In this paper, a novel method based on TOPSIS is proposed. TOPSIS combines four centralities to analyze the influence of nodes. Centralities combined in TOPSIS are Degree, betweenness, closeness, and eigenvector centrality. The top 10 lists of important nodes are compared between the proposed approach and the centralities cited above. Furthermore, to evaluate the effectiveness of the proposed methodology, the SI model is used to calculate the spreading effect of nodes using a single criterion such as DC, BC, CC, and EC, and multiple criteria in the TOPSIS-based approach. The experimental results on the Facebook ego network show that the TOPSIS-based approach has superior performance to that single criterion as shown in Fig. 5, TOPSIS plot is always superior to DC, BC, CC, and EC.

In this paper, the proposed approach is used with subjective weights but objective weights can enrich results. Therefore, a future research topic will focus on an integration method between TOPSIS and other objective weighting methods. Another future direction is to exploit these results to start a clustering algorithm for the purpose of community detection.

Abbreviations

MCDM	Multi criteria decision making
TOPSIS	Technique for order of preference by similarity to ideal solution
SI	Susceptible Infected

DC	Degree centrality
BC	Betweenness centrality
CC	Closeness centrality
EC	Eigenvector centrality

References

- Barabási, A.-L., Gulbahce, N., Loscalzo, J.: Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011). <https://doi.org/10.1038/nrg2918>
- Bonacich, P., Lloyd, P.: Eigenvector-like measures of centrality for asymmetric relations. *Soc. Netw.* **23**, 191–201 (2001). [https://doi.org/10.1016/S0378-8733\(01\)00038-7](https://doi.org/10.1016/S0378-8733(01)00038-7)
- Brandes, U.: A faster algorithm for betweenness centrality*. *J. Math. Sociol.* **25**, 163–177 (2001). <https://doi.org/10.1080/0022250X.2001.9990249>
- Brandes, U., Borgatti, S.P., Freeman, L.C.: Maintaining the duality of closeness and betweenness centrality. *Soc. Netw.* **44**, 153–159 (2016). <https://doi.org/10.1016/j.socnet.2015.08.003>
- Bródka, P., Skibicki, K., Kazienko, P., Musiał, K.: A degree centrality in multi-layered social network. In: 2011 International Conference on Computational Aspects of Social Networks (CASoN), pp. 237–242 (2011). <https://doi.org/10.1109/CASON.2011.6085951>
- Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198 (2009). <https://doi.org/10.1038/nrn2575>
- Chen, S.-J., Hwang, C.-L.: Fuzzy multiple attribute decision making methods. In: Chen, S.-J., Hwang, C.-L. (eds.) *Fuzzy Multiple Attribute Decision Making*, pp. 289–486. Springer, Berlin (1992). https://doi.org/10.1007/978-3-642-46768-4_5
- Cheng, S., YunTao, P., JunPeng, Y., Hong, G., ZhengLu, Y., ZhiYu, H.: PageRank, HITS and impact factor for journal ranking. In: 2009 WRI World Congress on Computer Science and Information Engineering. Presented at the 2009 WRI World Congress on Computer Science and Information Engineering, pp. 285–290 (2009). <https://doi.org/10.1109/CSIE.2009.351>
- Du, Y., Gao, C., Chen, X., Hu, Y., Sadiq, R., Deng, Y.: A new closeness centrality measure via effective distance in complex networks. *Chaos Interdiscip. J. Nonlinear Sci.* **25**, 33112 (2015)
- Galam, S., Zucker, J.-D.: From individual choice to group decision-making. *Phys. Stat. Mech. Its Appl.* **287**, 644–659 (2000). [https://doi.org/10.1016/S0378-4371\(00\)00399-X](https://doi.org/10.1016/S0378-4371(00)00399-X)
- Hamer, W.H.: *The Milroy Lectures on Epidemic Disease in England*. Bedford Press (1906)
- Katz, L.: A new status index derived from sociometric analysis. *Psychometrika* **18**, 39–43 (1953). <https://doi.org/10.1007/BF02289026>
- Leskovec, J., McAuley, J.: Learning to discover social circles in ego networks. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc. (2012)
- Mendes, G.A., da Silva, L.R., Herrmann, H.J.: Traffic gridlock on complex networks. *Phys. Stat. Mech. Appl.* **391**, 362–370 (2012). <https://doi.org/10.1016/j.physa.2011.07.046>
- Resende, M.G.C., Pardalos, P.M. (eds.): *Handbook of Optimization in Telecommunications*. Springer US, Boston, MA (2006). <https://doi.org/10.1007/978-0-387-30165-5>
- SNAP: Network datasets: Social circles [WWW Document] (2022). <https://snap.stanford.edu/data/ego-Facebook.html>. Accessed 11 May 2022
- Xiong, X., Hu, Y.: Research on the dynamics of opinion spread based on social network services. *Wuli XuebaoActa Phys. Sin.* **61** (2012)



Evaluating FFT-Based Convolutions on Skin Diseases Dataset

Amina Aboulmira¹(✉), Hamid Hrimech¹, and Mohamed Lachgar²

¹ Hassan 1er University, ENSA, Berrechid, Morocco
a.aboulmira@uhp.ac.ma

² Chouaib Doukkali University, ENSA, El Jadida, Morocco

Abstract. Deep Learning has been used in many challenges in the current Big Data era, with great improvements, especially in image classification tasks, helping to boost the potential efficacy in computer vision and signal processing applications. As deep learning techniques develop more and more, automated classification tasks based on convolutional neural networks have integrated many fields and have become indispensable nowadays. Over the past decade, these architectures have become very sophisticated. Although several strategies for deep structures have been used, the frequency domain has not been widely investigated in this field. This research suggests using Fast Fourier convolution layers to perform image classification for skin disorders with a denoising block. The obtained results show that the CNN approach with the FFT considerably decreases the training time compared to a naïve CNN model with convolution. The suggested method was tested using the publicly accessible HAM10000 dataset and achieved an accuracy rate of 88%.

Keywords: Skin lesion · Classification · Fast Fourier · Frequency domain · Convolutional neural networks

1 Introduction

The appearance of ANNs dates back to the 1950s, but due to the major limitations at the time (the overfitting problems that hindered the learning of the deep architecture, the limited computational resources and mainly the lack of sufficient data), the technology has not been very successful in real applications [1]. However, Deep Learning today has become indispensable in many fields and has attracted a lot of attention. Thanks to its performance and the development of several techniques to target different problems in various fields [2]. Indeed, these deep learning techniques have become popular in the medical field thanks to their performance in imitating the human performance in the realization of certain tasks such as the detection and classification of diseases.

Skin conditions can cause internal complications, as well as the spread of the disease from one person to another, if not treated at an early stage. Therefore, by examining the affected area earlier, skin infections can be prevented. Images of skin infections have helped in the development of several effective and reliable approaches to automatically

detect skin diseases and limit their severity. As a result, early diagnosis of these diseases is a primary task and plays a crucial role in reducing mortality. Over the years, many researchers have embarked on this adventure, exploiting various artificial intelligence techniques to locate skin lesions at an early stage to prevent the development of malignant tumors and other severe infections [3].

In the literature, different approaches have been proposed for the classification of skin diseases using medical images. Most of the proposed works usually try to detect the features of these medical images on their raw pixels. However, learning and analyzing the frequencies present in the images can be useful, as the frequency information can greatly help in discriminating between textures [4].

To train automatic algorithms that can distinguish these skin lesions, which are difficult to diagnose due to the millimeter differences between these tumors and infections, a robust and deep model has to be developed. Fast Fourier convolution is a powerful technique that has proven to be a better alternative to convolution in terms of complexity and computational time, due to its ability to perform convolution with the same efficiency using pointwise multiplication in the Fourier domain. This significantly reduces the learning time, as it has been tested and experimented with in several types of research [5]. However, when the images are convolved with small filter kernels, the learning time becomes slower [6]. The interest of using the FFT in our research is not to reduce the training time, but to exploit the frequency domain in the preprocessing and feature extraction step by analyzing the low-level and high-level features of these images, and see the impact of frequencies on the model performance.

The main contributions of this paper are:

- Implementation of a fft-based convolution layer for ResNet-18.
- Analyzing how the features extracted by a frequency domain network can be exploited.
- Evaluation of the frequency domain in skin disease dataset.

The remainder of this paper is organized as follows: The following section presents a brief introduction. The related work is discussed in Sect. 2. Section 3 exposes a background of the Fast Fourier Transform and convolutions. The methodology is presented in Sect. 4, particular the public skin disease datasets description, the preprocessing stage and the conception of the FFT convolutional process. The Sect. 5 present simulation results and analyse these results. The conclusion and future works are finally presented in the Sect. 6.

2 Related Works

The frequency domain is very popular in signal processing audio classification and a little less in image processing [7]. However, the use of the frequency domain to improve the accuracy of image classification is not yet widely exploited. In this context, this study provides a Fourier layer that perform the operation of convolution between kernel and input image with the introduction of a denoising block in frequency domain.

In the literature, studies investigate the frequency domain in different area. In [8], Xu et al. proposed process images using the Discrete Cosine Transform (DCT) by computing

the spectrogram of each image. They also selected only the 5 highest frequency components as input to the model instead of using the whole image. The proposed approach achieves an improvement of 1.60% in the accuracy rate on the Imagenet dataset.

Qin et al. proposed a frequency channel attention network to exploit the information of the different frequency components of the DCT, adding a channel attention that guides the selection of channel features [9].

Stuchi et al. in [10], designed a set of frequency domain filters with fixed parameters to operate on the images and extract the discriminative feature using Fourier analysis. This improves the detection of face liveliness.

Wang et al. in [11] analyzed the effect of high frequency components on DNNs and noted that these frequencies can affect the perception of the pattern. DNNs are adapted to understand high frequencies that are almost imperceptible to humans.

In [12], Cai et al., worked on image restoration and synthesis in the frequency domain, they showed that Fourier transformation with FFT is a good techniques for extracting high-frequency information which enhances better the identity preserving in the image generation process.

In the classification of skin diseases, several artificial intelligence algorithms have been exploited in the different proposed researches, however the frequency domain is not really invested. The proposed approaches for the classification of the skin diseases, tried to use several deepLearning techniques to train efficient classifiers.

Popescu et al. [13] presented a system based on a multi-network ensemble. They combined nine neural networks for classification of the seven classes presented in the HAM10000 dataset. Weights were assigned to each network, as well as to the predicted class within each network. The final decision is made by merging decisions considering all weights. The approach had a significant increase in accuracy, which increased by about 3% compared to the performance of the best individual network.

By combining handcrafted features of the ABCD medical algorithm (asymmetry, border, color and diameter) with deep learning features and using mutual information measures, Almaraz-Damien et al. [14] achieved an overall accuracy rate of 90.67% in the classification of melanoma and nevus. The proposed system was based on the concept of mutual information and several CNN architectures were proposed for data extraction, which will be merged with the handcrafted features. The classification was done by several classifiers, including linear regression (LR), support vector machines (SVM) and relevant vector machines (RVM).

Kang et al. [15] also exploited deep learning to classify skin diseases by combining the MobileNet CNN, known by a reduced number of parameters and long-term memory (LSTM). The approach was evaluated on the HAM1000 dataset and showed a significant performance of 85%. MobileNetV2 performed well compared to other classifiers in this study, which proved to be very advantageous for mobile system design, since this model guarantees reduced computational expenses and compatibility with mobile devices.

3 Background

3.1 Image Denoising

Image denoising plays a critical role in medical image analysis where noise is present such as random variation in brightness or color intensities in the images, which can be produced by the image sensor or circuitry of a scanner [16]. It helps to eliminate noise in the image, which improves image quality. Indeed, low image quality is an obstacle to feature extraction and image analysis. Therefore, these noises must be removed to ensure better understanding of medical images to detect tumors, cancers and other diseases.

The denoising phase is established in the frequency domain, where the fft is applied to the input image, then the frequencies associated with each coefficient are calculated and only the coefficients that have a sufficiently low frequency (in absolute) are kept.

3.2 FFT-Based Convolution

The fast Fourier transform is a technique known as one of the greatest developments in computer science, many FFT algorithms have been proposed for various fields. In image processing, the Fast Fourier Transform (FFT) is one of the most important computer developments, it is an algorithm to convert spatial functions (such as images) into a representation in the frequency domain [17].

The application of the Fourier transform can be used in various deep learning algorithms and has many advantages for convolutional neural networks. Indeed, the Fourier transform allows to simulate convolution in the spatial domain by transforming the computation of layers (sliding window operation) into element-by-element products in the frequency domain, while keeping the same performance [18]. Letting $(*)$ denote the convolution and $(\circ)$ an element-wise multiplication, the algorithm is summarized by Eq. 1, where F is Fourier transform and F^{-1} its inverse. The convolutions between I and K can be computed by:

$$I * K = F^{-1}(F(I) \circ F(K)) \quad (1)$$

Inspired by the work of Mathieu and LeCun [19], who used the FFT to accelerate the training of convolutional neural networks in a graphics processing unit (GPU) architecture, FFT-based convolution block was performed.

The input image is transformed into the frequency domain using the FFT, it then goes through a denoising phase where the noise due to the different modes of image acquisition is reduced. The transformed image is then multiplied with the transformed filter also in the frequency domain, then transformed back into the time domain using the inverse DFT. The generated feature maps are obtained by multiplying the two signals that are convolved. More specifically, the real and imaginary parts of the input image are multiplied by the real and imaginary parts of the kernel.

The process of the algorithm is completed by returning to the spatial domain using the inverse FFT (IFFT), the final convolved image is obtained. However, before proceeding, the input image $(N * N)$ and the kernel $(K * K)$ must have the same size after filling. However, before proceeding, the input image $(N * N)$ and the kernel $(K * K)$ must have

the same size after filling. Therefore, the image and kernel are filled with zeros of length $(N + K - 1)$, this avoids the aliasing effect. Once the convolution is done, the FFT inverse is used to return to the spatial domain and the extra values are removed. Figure 1 shows the process of the approach.

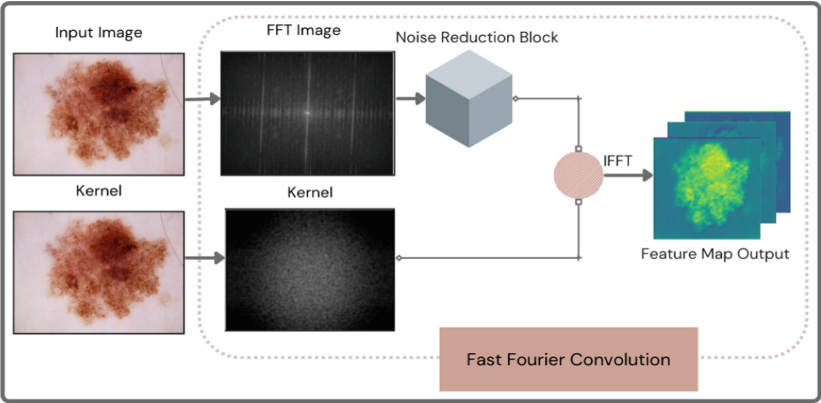


Fig. 1. The fft-based convolution block.

4 Methodology

4.1 Workflow

Figure 2 illustrates the overall skin disease classification process. The basic architecture is Resnet-18, where the convolutional layers have been replaced by our fft-based convolutional layer. The pre-trained weights from Resnet-18 by Imagenet are then added to this new architecture, which retains the same shape as the previous one. Training is finally performed using transfer learning.

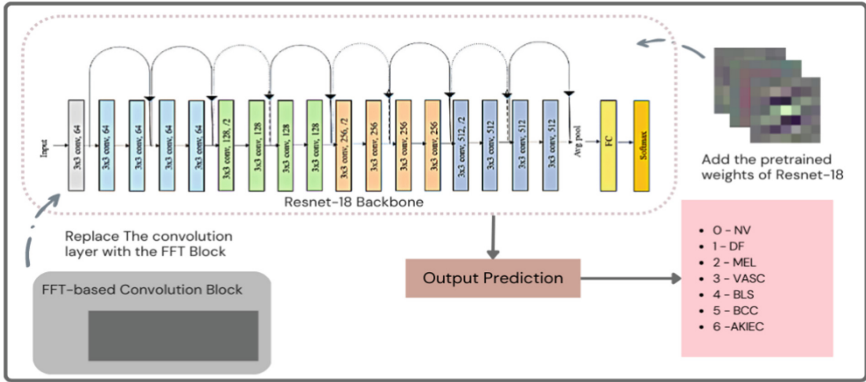


Fig. 2. General workflow of our process.

4.2 Dataset

HAM10000 (Human Against Machine) is a dermoscopy image dataset which is comprised of 10015 dermatoscopic images from patients of various ages and genders. It represents the most common skin cancers divided into seven classes: actinic keratosis (AKIEC), basal cell carcinoma (BCC), benign keratosis (BKL), dermatofibroma (DF), melanoma (Mel), melanocytic nevus (NV) and vascular lesions (VASC).

In this experiment, our dataset was randomly divided into a training, validation and test set in a 70:20:10.

4.3 Data Augmentation

Data augmentation aims to increase the amount of data by performing transformations on the already existing data in the dataset. Different transformations were performed, such as random horizontal and vertical flipping, random rotation and color normalization.

Indeed, the variance in skin tone between individuals, the distribution of color values varied by lighting conditions, cameras, and other factors, as noted by Kawahara et al.20 has a negative impact on the results and is therefore characterized as noise. For this reason, a color normalization phase was applied by averaging and normalizing three channels over the entire data set.

4.4 Resnet-18

Resnet was proposed by Kaiming He et al. in 2016 [20] and proved to be very successful, winning an accuracy rate of about 80% on the ILSVR Challenge. Each layer of the network learns residual characteristics with reference to the layer's inputs, according to the concept of the residual layers. There are many architecture variants of ResNet. ResNet-18 architecture was used in this study, it represents a good compromise between increasing depth (18 deep layers) and gaining performance.

4.5 Transfer Learning

The main goal of transfer learning (TL) is to use the knowledge gained from another dataset to another similar particular domain [21]. The pre-trained weights on Imagenet Dataset were used to train our dataset, which speeds up the training phase instead of starting from scratch.

Thus, ResNet-18 will act as our feature extractor. Finally, a fully connected layer is added for the classification of the seven classes of the HAM10000 dataset.

5 Experimental Results and Discussion

To evaluate the performance of the proposed methodology, the modified Resnet-18 is trained on the HAM10000 dataset for the classification of the 7 common skin diseases. The method is implemented in PyTorch [22] and the experiments were done with GeForce RTX 2060 GPU – 6 GB GDDR5 memory and a memory (RAM) of 16.0GB on a Windows10 machine. The experimental results in Table 1 summarizes the classification report on testing dataset for the different networks.

Due to resource limitations, small input images with size of 64 * 64 are used, with a batch size of 16, Adam optimizer with a learning rate of 0.001, the categorical cross entropy loss function and 15 epochs. Figure 3 shows training and validation accuracy curves for the proposed method with the FFT-based convolution and the denoising phase.

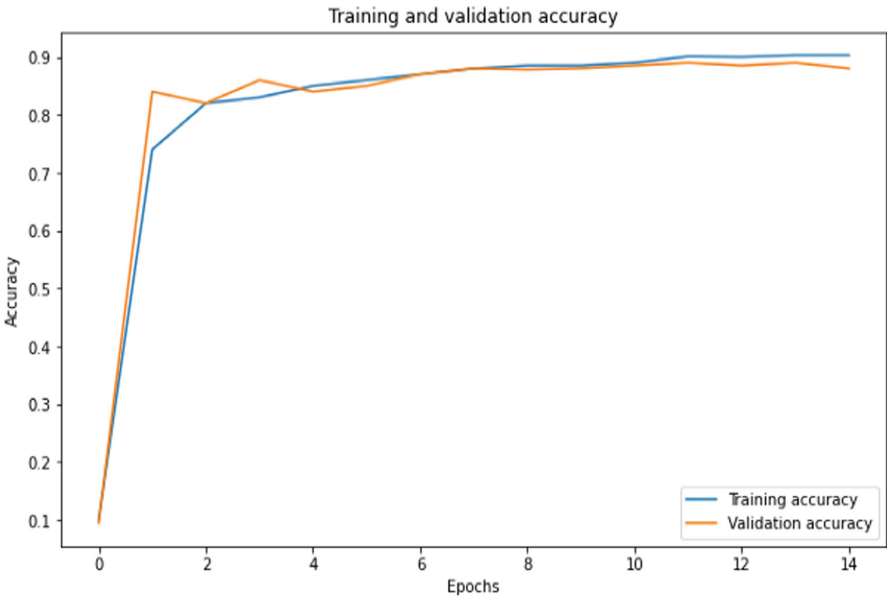
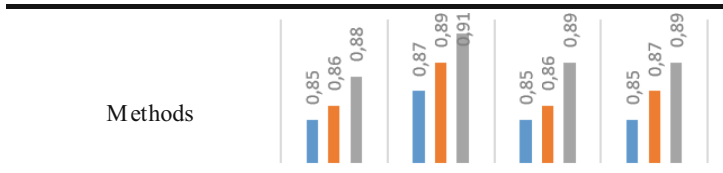


Fig. 3. Training and validation accuracy for our proposed method

It can be observed that the training process was carried out without overfitting. The validation accuracy curve simultaneously follows the performance values of the training.

As the number of epochs increases, the performance of the model increases, until epoch 13 where there is no more improvement in accuracy, which means that the model has reached its maximum performance. The resulting accuracy was 90.3%. The results of the training and validation show that the model was able to have a good generalization, this is promising for future works, where images of larger size can be used. The average values of the experimental results are shown in Table 1.

Table 1. Comparison of overall performance: accuracy, precision, recall and f-1 score.



Methods	Accuracy	Precision	Recall	F1-score
Resnet-18	0.85	0.87	0.85	0.85
Resnet-18 with FFT	0.86	0.89	0.86	0.87
Resnet-18 with FFT and denoising Block	0.88	0.91	0.89	0.89

The results with FFT-based convolution are much higher than direct convolution, even if the two models had the same architecture, and the same optimizer and loss function. The third method based on FFT convolution and block reduction outperformed the other methods. This led us to the capability of Fast Fourier technique and the denoising phase to extract more features than a convolution operation, where the frequencies can be exploited in various ways to enhance accuracy classification like detecting edges or reducing noises. It should be noted that in frequency domain, there is high frequencies which represent fine details like borders, edges and noise, and low frequencies which present a large scale information, such as shape and homogeneous areas [10].

However, it is essential to note that the method based on the Fast Fourier transform takes more time than the direct convolution in our experiment. This may seem contradictory since the Fourier transform reduces the complexity of functions. But, this is only valid for large filter kernels (larger than about 10×10 pixels). The execution time is therefore proportional to the length of the kernel [23].

In addition, when the filters and images are of different sizes, the calculation of FFT coefficients requires additional intermediate memory [24]. However, due to limited memory resources, the intermediate memory accumulation of the FFT algorithm has been an obstacle. Indeed, this led us to reduce the size of the images, as well as the batch-size, which affect network performance.

6 Conclusion

A Fourier algorithm that simulates the convolutional layer with the implementation of the Fast Fourier Transform is presented. It outperforms the direct convolution performed

by Pytorch in terms of accuracy, as verified by numerical experiments. The effectiveness of this approach for skin disease classification has thus been proven, opening up many possibilities for taking advantage of the frequency domain with CNNs for skin disease classification.

However, the performance of the model can be improved if we further investigate frequency learning and overcome the computational limitations for better recognition accuracy. In future work, we intend to implement an FFT-based feature extractor with large kernels to learn in the frequency domain layers to better understand the effect of these frequencies and how we can explore them to have a significant impact on improving CNN model performance and reducing computation time.

Abbreviations

CNN	Convolutional neural network
ANN	Artificial neural network
DNN	Deep Neural Network
FFT	Fast Fourier Transform
IFFT	Inverse fast Fourier transform
HAM1000	Human Against Machine

References

1. Lee, J.-G., et al.: Deep learning in medical imaging: general overview. *Korean J. Radiol.* **18**(4), 570–584 (2017). <https://doi.org/10.3348/kjr.2017.18.4.570>
2. Farsal, W., Anter, S., Ramdani, M.: Deep learning: an overview. In: *Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications*, pp. 1–6. New York, NY, USA (2018). <https://doi.org/10.1145/3289402.3289538>
3. Abunadi, I., Senan, E.: deep learning and machine learning techniques of diagnosis dermoscopy images for early detection of skin diseases. *Electronics* **10**, 3158 (2021). <https://doi.org/10.3390/electronics10243158>
4. Saldanha, N., Pinteá, S.L., van Gemert, J.C., Tomen, N.: Frequency learning for structured CNN filters with Gaussian fractional derivatives (2021). <https://doi.org/10.48550/arXiv.2111.06660>
5. Vasilache, N., Johnson, J., Mathieu, M., Chintala, S., Piantino, S., LeCun, Y.: Fast Convolutional Nets With ffbft: A GPU Performance Evaluation (2014). <https://doi.org/10.48550/arXiv.1412.7580>
6. Lavin, A., S. Gray, S.: Fast algorithms for convolutional neural networks. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4013–4021. Las Vegas, NV, USA (2016). <https://doi.org/10.1109/CVPR.2016.435>
7. Adachi, F., Tomeba, H., Takeda, K.: Introduction of frequency-domain signal processing to broadband single-carrier transmissions in a wireless channel. *IEICE Trans. Commun.* **E92-B**(9), 2789–2808 (2009)
8. Xu, K., Qin, M., Sun, F., Wang, Y., Chen, Y.-K., Ren, F.: Learning in the Frequency Domain (2020). <https://doi.org/10.48550/arXiv.2002.12416>
9. Qin, Z., Zhang, P., Wu, F., Li, X.: FcaNet: Frequency Channel Attention Networks (2020). <https://doi.org/10.48550/arXiv.2012.11879>

10. Stuchi, J., et al.: Improving image classification with frequency domain layers for feature extraction, pp. 1–6 (2017). <https://doi.org/10.1109/MLSP.2017.8168168>
11. Wang, H., Wu, X., Huang, Z., Xing, E.: High-Frequency Component Helps Explain the Generalization of Convolutional Neural Networks, pp. 8681–8691 (2020). <https://doi.org/10.1109/CVPR42600.2020.00871>
12. Cai, M., Li, Y.: Out-of-distribution Detection via Frequency-regularized Generative Models ArXiv (2022). <https://doi.org/10.48550/arXiv.2208.09083>
13. Popescu, D., El-khatib, M., Ichim, L.: Skin lesion classification using collective intelligence of multiple neural networks. *Sensors* **22**, 4399 (2022). <https://doi.org/10.3390/s22124399>
14. Almaraz Damian, J.A., Ponomaryov, V., Sadovnychiy, S., Castillejos-Fernandez, H.: Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures. *Entropy* **22**, 484 (2020). <https://doi.org/10.3390/e22040484>
15. Kang, J.J., Srinivasu, P., Gnana Siva Sai, J., Ijaz, M.F., Bhoi, A.K., Kim, W.: Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors* **21**, 2852 (2021). <https://doi.org/10.3390/s21082852>
16. Mohd Sagheer, S.V., George, S.N.: A review on medical image denoising algorithms. *Biomed. Signal Process. Control* **61**, 102036 (2020). <https://doi.org/10.1016/j.bspc.2020.102036>
17. Buijs, H., Pomerleau, A., Fournier, M., Tam, W.: Implementation of a fast Fourier transform (FFT) for image processing applications. *IEEE Trans. Acoust. Speech Signal Process.* **22**(6), 420–424 (1974). <https://doi.org/10.1109/TASSP.1974.1162620>
18. Chi, L., Jiang, B., Mu, Y.: Fast fourier convolution. In: *Advances in Neural Information Processing Systems*, vol. 33, pp. 4479–4488 (2020). Accessed 19 Oct 2022
19. M. Mathieu, M. Henaff, and Y. LeCun, “Fast Training of Convolutional Networks through FFTs.” arXiv, Mar. 06, 2014. doi: <https://doi.org/10.48550/arXiv.1312.5851>
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
21. A survey of transfer learning | Journal of Big Data | Full Text. <https://journalofbigdata.springeropen.com/articles/https://doi.org/10.1186/s40537-016-0043-6> Accessed 19 Oct 2022
22. PyTorch. <https://pytorch.org/>. Accessed 19 Oct 2022
23. A Closer Look at Image Convolution. <http://www.dspguide.com/ch24/7.htm>. Accessed 19 Oct 2022
24. Abtahi, T., Kulkarni, A., Mohsenin, T.: Accelerating convolutional neural network with FFT on tiny cores, pp. 1–4 (2017). <https://doi.org/10.1109/ISCAS.2017.8050588>



On the Detection of Anomalous or Out-of-Distribution Data in Vision Models Using Statistical Techniques

Laura O'Mahony^(✉), David JP O'Sullivan, and Nikola S. Nikolov

University of Limerick, Castletroy V94 T9PX, Limerick, Ireland
{lauraa.omahony,david.osullivan,nikola.nikolov}@ul.ie

Abstract. Out-of-distribution data and anomalous inputs are vulnerabilities of machine learning systems today, often causing systems to make incorrect predictions. The diverse range of data on which these models are used makes detecting atypical inputs a difficult and important task. We assess a tool, Benford's law, as a method used to quantify the difference between real and corrupted inputs. We believe that in many settings, it could function as a filter for anomalous data points and for signalling out-of-distribution data. We hope to open a discussion on these applications and further areas where this technique is underexplored.

Keywords: Anomaly detection · Out-of-distribution data · Computer vision · Benford's law · Image corruption

1 Introduction

Deploying machine learning systems in the real world brings with it multiple AI safety challenges, for example, issues related to model robustness and monitoring. Research on the detection of adversarial examples, anomalous inputs, and flagging when the data a model is deployed on is out-of-distribution attempts to tackle such issues. Out-of-distribution means there is a difference in distributional properties between the test, training, and real world data. For machine learning models, even simple shifts in data distribution can lead to a large drop in performance [20]. An example of a failure caused by data anomalies is the discovery that classifiers often fail by giving high-confidence predictions on anomalous inputs when their predictions are incorrect [25]. The failure of such classifiers to indicate where they are likely wrong is problematic [4]. In this work, we study how a range of image corruptions appreciably deviate in distribution from natural images by leveraging a dataset containing various image corruption types. Our work trials the broad applicability of a method that could act as a filter for anomalous inputs and indicate distribution shifts. We find the method reflects some corruption types in the experiments performed and thus may be practical as an accompaniment to other methods for out-of-distribution data and anomalous input detection.

A vast amount of literature exists aiming to detect out-of-distribution and anomalous data [37]. Some of these methods focus on analysing the activations of hidden layers [14, 18, 26, 35]. Many works formulate the problem as a model

selection between an in-distribution and an out-of-distribution density model and use a likelihood-ratio based approach to detect out-of-distribution data [7, 33, 38]. Another approach relates to the softmax distribution [16]. Other authors utilise early-layer outputs [1] or gradient space in out-of-distribution data detection [19]. A technique known as outlier exposure tackles classifier overconfidence by training anomaly detectors against an auxiliary outlier dataset [17]. There also exists previous literature concerning the properties, such as the mean, of a dataset computed in the presence of outliers [36]. Other work on the low-order statistics of data approaches the problem of out-of-distribution data by analysing the low-order statistics and framing it as a statistical hypothesis testing problem [13].

Our approach differs in that it uses a less-known property of image distributions relating to Benford’s law (BFL), which has a proven track record in image forensics, as is discussed in Sect. 2. To our knowledge, the use of Benford’s law is novel in the context of out-of-distribution and anomaly detection in the context of machine learning classification. We note that our approach is compatible with other differing out-of-distribution heuristics and anomaly detection techniques, meaning it could complement other methods in the literature. Furthermore, our approach is surprisingly simple and computationally inexpensive as it does not modify the model parameters or architecture, making it scalable.

The paper is structured as follows: In Sect. 2, we outline the theory and current applications for the distributional summaries of images. In Sect. 3, we outline the pipeline used to measure an image’s deviation from its theoretical distribution before introducing the dataset used to evaluate this pipeline. We then present the experimental results. We also discuss possible extensions to this work and its applications before concluding in Sect. 4.

2 Background

In this section, we introduce the Benford equation as a benchmark that can be used to compare the statistical distributional properties of images. In 1938, Benford [6] introduced BFL for numbers. This paper predicts the statistical frequency of the leading digits (LDs) for a set of measurements of a broad range of “natural” quantities [29]. Here, natural data is understood to be data collected by observing natural phenomena. It is worth noting that the cases where the rule applies are often determined experimentally rather than mathematically [10]. Benford’s law has been empirically observed for many natural quantities, such as the lengths of rivers, population numbers, stock market prices, etc. The predicted distribution, $p(d)$ is given by Eq. 1, where d is the LD of the number (e.g., ‘245326’ has the LD ‘2’. The remaining digits ‘45326’ are discarded as Benford’s law only concerns the distribution of the LDs.),

$$p(d) = \log_b \left(1 + \frac{1}{d} \right), \quad d = 1, 2, \dots, b - 1. \quad (1)$$

As described in the equation, the nature of probabilities of the first digits is logarithmic to the base b , for invariant b . The distribution given by the Benford equation, i.e., Eq. 1, in base 10, is shown in Fig. 1. As shown in this figure, the occurrence of the smaller first digits is more frequent than larger first digits.

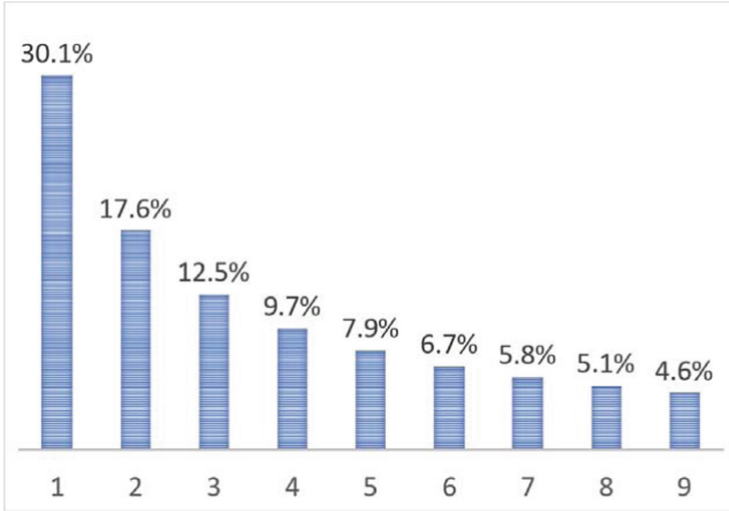


Fig. 1. The theoretical distribution of first digits in data that obeys Benford’s law. The height of the bar is the percentage of numbers that start with the digit.

It has been observed that this rule is not well-fitted by the LD statistics of altered data [11]. As a result, it has been useful for purposes such as fraud detection in taxing and accounting [5,12]. It has also been used to detect anomalies that require investigation for fraud in customs declarations [9].

BFL can also be applied to image data, provided an appropriate transform is made before the distribution is extracted from images. The discrete cosine transform (DCT) [3] is a widely used transformation technique in signal processing and data compression [30]. It expresses a finite sequence of data points as a sum of cosine functions oscillating at various frequencies. It has been shown that the DCT coefficients of “natural” images can typically be modelled by a Laplacian-like distribution, which follows BFL [22]. Therefore, images mapped into the DCT domain typically follow BFL closely [28], giving us a transform that can be used to extract a distribution that should follow BFL. Previous work has exploited this for applications such as detecting JPEG compression [24,27], synthetic imagery [2], estimating the amount of processing that has been applied to an image, and detecting image tampering [34]. It has been applied to detect if an image is generated by a generative adversarial network (GAN) [8]. It has also been shown that images that reflect only a small part of a scene do not fit BFL well [2]. A connection between a neural network’s generalisation capability and a BFL related metric has been studied [32].

The extensive analysis of BFL in image forensics and otherwise lends the question of whether the rule could be employed in AI safety applications, such as detecting anomalies and indicating if data is out-of-distribution. We ask if this theoretical structure of natural images shifts as images become less legible, allowing us to detect anomalies and reflect when data is out-of-distribution.

3 Analysis of Image Discrete Cosine Transform Coefficients

3.1 Methodology for Extracting a Distributional Comparison Measure from Images

In this section, we provide the details of how images are processed to analyse the DCT coefficient distribution of images and how we measure the divergence of the empirical distribution of images from the BFL theoretical distribution.

We first need to obtain a distributional representation of an image so that we can compare it to the theoretical distribution. The process of obtaining this distribution and measuring the divergence from BFL is depicted in Fig. 2. Given an RGB image I , the image is initially converted to greyscale. The image is then divided into 8×8 pixel blocks that do not overlap. Each block is transformed using the DCT. The empirical distribution of the LDs of the DCT coefficients from each block is calculated with respect to a base, e.g., base 10. This outputs a probability mass function (pmf), $\hat{p}(d)$, for the image in question, I . This computed pmf is compared to the theoretical distribution, given by the Benford equation, i.e., Eq. 1. A statistical metric, the Jensen-Shannon divergence [23], is employed to investigate the difference between these distributions. This gives us D_{JS} , a measure of the distributional difference between the empirical and theoretical distribution of I .

This process provides a measure of how far the distribution of an image is from BFL. We can then use this as a comparison between clean images and images with different corruption types applied at varying severity levels, which we could think of as distributional shifts in the data. To test this process, we use the Image-C dataset [15], which we describe in the next section.

3.2 Data

To test our pipeline described in Sect. 3.1, the ImageNet-C dataset was repurposed to assess whether varying levels of image corruption are reflected by deviations of images from BFL.

ImageNet-C¹ [15] is a dataset created by Hendrycks and Dietterich, which is designed to establish benchmarks for image classifier robustness. The dataset contains 15 distinct types of image corruptions (noise, blur, weather, and digital corruptions). Each of these corruption types has severity levels 1 – 5, where

¹ License [Creative Commons Attribution 4.0 Intern](https://creativecommons.org/licenses/by/4.0/), DOI: 10.5281/zenodo.2235448ational.

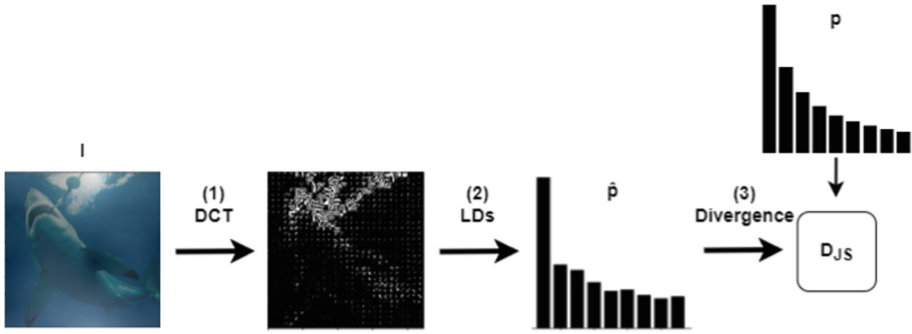


Fig. 2. The procedure of quantifying the difference between the distribution of the image discrete cosine transform (DCT) coefficients and the theoretical distribution given by Benford’s law consists of (1) applying a discrete cosine transform to the image I , followed by (2) calculating the leading digit (LD) distribution of the transform coefficients, and finally (3) a statistical comparison between the theoretical distribution p and observed distribution $\hat{p}$ of the transformed coefficients is made. The raw image I is taken from the ImageNet-C dataset [15].

higher levels of severity indicate a higher level of corruption being applied. The dataset is algorithmically generated from the ImageNet validation data [31] and intended to evaluate the performance of a model on common image corruptions. Therefore, the data is the standard ImageNet size and consists of 1000 ImageNet classes. Both the type and severity of corruption give us a unique dataset to test if we can detect differences in the distribution of images using our method. See Fig. 3 for an example of an image with a subset of the corruption types applied at corruption severity levels 1, 3 and 5 respectively. Four different corruption types are shown, corresponding to a common image from the ImageNet validation data.

3.3 Results of the Image Comparison Metric and Discussion

The pipeline detailed in Sect. 3.1 was applied to the 50,000 ImageNet validation dataset [31] images. Due to computational constraints, it was applied to the ImageNet-C images of corruption severity levels 1, 3 and 5, corruption types glass blur, contrast, Gaussian noise, and fog². We first look at the distribution of these divergence statistics. From Fig. 4, we can see that, typically, across all corruption types, with the exception of Gaussian noise, corrupted images show an increase in the median divergence from the Benford’s law (BFL) benchmark compared to uncorrupted images. Higher severity levels of corruption types glass blur and contrast lead to a larger median divergence from the BFL benchmark compared to lower severity levels of corruption. A different pattern is seen for the Gaussian noise corrupted images at severity levels 1 and 3. A possible explanation

² Code available at <https://github.com/lomahony/first-digits>.

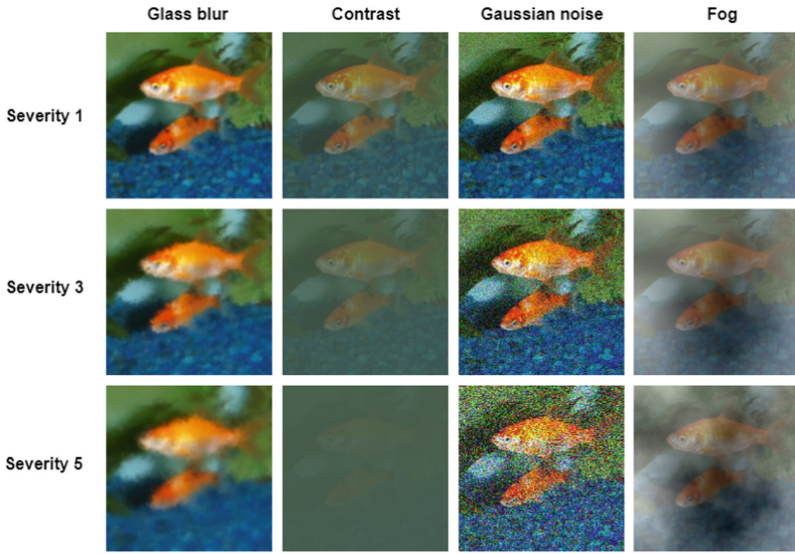


Fig. 3. An example of images taken from ImageNet-C [15] representing different corruption types.

is that Gaussian noise may interfere with the Laplacian-like distribution followed by the DCT coefficients of images.

Further experiments assessed whether the proposed approach can capture model degradation on out-of-distribution data. A popular pretrained image classification model called AlexNet [21] was used as a benchmark model to compare the test accuracy of the top-1 prediction accuracy at the selected corruption levels. Figure 5 shows that, as expected, the baseline ImageNet validation accuracy of AlexNet of 0.633 decreased as the level of corruption increased for each corruption category. In some cases, the model performance drop did not reflect a higher median divergence from the BFL benchmark, such as the median Jensen-Shannon divergence for Gaussian noise corrupted images at severity levels 1 and 3 compared with the median divergence for the clean ImageNet images. In other cases, a considerable jump in the median divergence from the BFL benchmark reflected a large model performance degradation. An example of this is the jump in Jensen-Shannon divergence between contrast and Gaussian noise corrupted images at severity levels 3 and 5. This was reflected by a very low top-1 accuracy. From these results, a clear limitation of this technique is that the distributional shift may impact model performance significantly more than what is reflected by the divergence statistic. This suggests this approach may be useful as a complementary approach but should not replace other anomaly or out-of-distribution detection methods. Another limitation of this approach is that it may be useful for sensory anomaly detection (detecting changes in the input space e.g., indicating where samples may be from a different domain), but not semantic anomaly

detection (detecting changes in the label space e.g., detecting the occurrence of new classes) [37].

A generalisation of BFL has been shown to better fit natural images due to the properties of the DCT transform [28]. This suggests an extension of this work could involve investigating if a generalised Benford's equation improves on the results presented here for Gaussian noise as well as the other corruption types. In addition, further extensions of this work could involve calculating the empirical distribution of training images. This could be used as the benchmark instead of Benford's equation to test if this technique could be used outside of natural images. As BFL is applicable to various data types, another potential future direction could involve testing the methodology on other data types. Further experiments could also involve testing to what extent model performance drop correlates with the level of distribution shift detected using our pipeline.

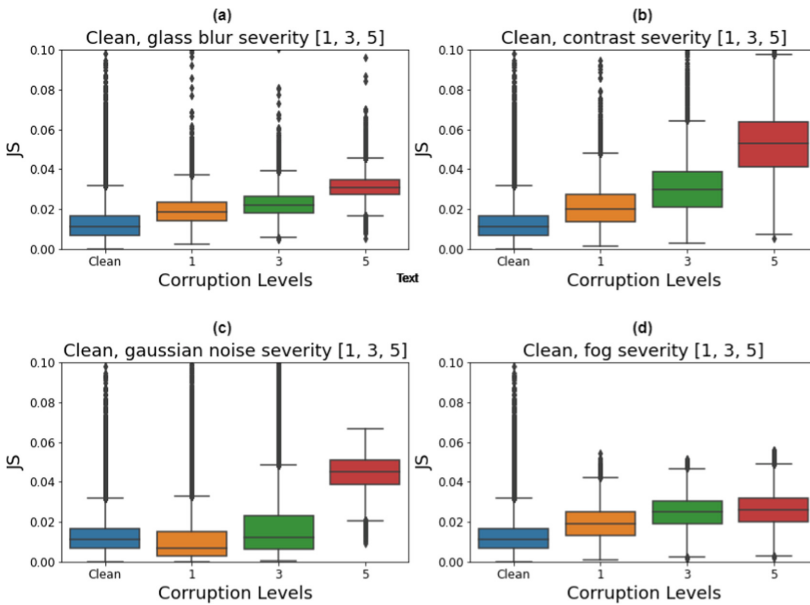


Fig. 4. The Jensen-Shannon divergence of the statistics of the discrete cosine transform coefficients for clean ImageNet images (blue), images of corruption severity 1 (orange), corruption severity 3 (green), and corruption severity 5 (red).

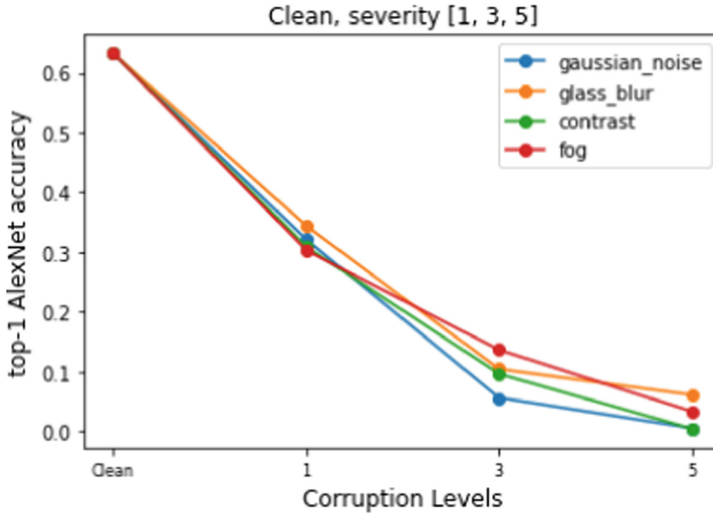


Fig. 5. The top-1 predictive accuracy evaluated using AlexNet [21] is shown for clean ImageNet images, images of corruption severity 1, corruption severity 3, and corruption severity 5.

4 Conclusion

In this paper, we investigated if the divergence of image statistics from Benford’s Law is reflected when images are corrupted at various levels. Importantly, we do not claim that this phenomenon is universal, but we found there to be a difference in image distribution across many corruption types for natural images taken from ImageNet. Results show that for many corruption types, images that are corrupted to a higher level typically deviate from the expected distribution more. This positive result motivates further testing, and we believe extensions of the methods employed in this paper are likely to give positive results. Future work might involve testing the extent that the proposed approach reflects the model degradation on out-of-distribution data. This technique could be added to the toolkit as a low computational filter for anomalous or out-of-distribution data or to assess the reliability of images before attempting to classify them. It could even be investigated as a coarse filter to detect some types of adversarial examples. Furthermore, wider ramifications of the successful differentiation in distribution between clean and corrupted images that the authors are not aware of may lead to further interesting research questions.

Acknowledgements. This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant Number 18/CRT/6049. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

1. Abdelzad, V., Czarnecki, K., Salay, R., Denouden, T., Vernekar, S., Phan, B.: Detecting out-of-distribution inputs in deep neural networks using an early-layer output. arXiv preprint [arXiv:1910.10307](https://arxiv.org/abs/1910.10307) (2019)
2. Acebo, E., Sbert, M.: Benford's law for natural and synthetic images. In: Proceedings of the 1st Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging, pp. 169–176 (2005)
3. Ahmed, N., Natarajan, T., Rao, K.R.: Discrete cosine transform. *IEEE Trans. Comput.* **100**(1), 90–93 (1974)
4. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., Mané, D.: Concrete problems in AI safety. arXiv preprint [arXiv:1606.06565](https://arxiv.org/abs/1606.06565) (2016)
5. Asllani, A., Naco, M., et al.: Using benford's law for fraud detection in accounting practices. *J. Soc. Sci. Stud.* **2**(1), 129–143 (2015)
6. Benford, F.: The law of anomalous numbers. In: Proceedings of the American Philosophical Society, pp. 551–572 (1938)
7. Bishop, C.M.: Novelty detection and neural network validation. *IEE Proc. Vis. Image Signal Process.* **141**(4), 217–222 (1994)
8. Bonettini, N., Bestagini, P., Milani, S., Tubaro, S.: On the use of benford's law to detect GAN-generated images. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 5495–5502. IEEE (2021)
9. Cerioli, A., Barabesi, L., Cerasa, A., Menegatti, M., Perrotta, D.: Newcomb-benford law and the detection of frauds in international trade. *Proc. Natl. Acad. Sci.* **116**(1), 106–115 (2019)
10. Crişan, D., et al.: Analyzing benford's law's powerful applications in image forensics. *Appl. Sci.* **11**(23), 11482 (2021)
11. Diekmann, A.: Not the first digit! using benford's law to detect fraudulent scientific data. *J. Appl. Stat.* **34**(3), 321–329 (2007)
12. Durtschi, C., Hillison, W., Pacini, C.: The effective use of benford's law to assist in detecting fraud in accounting data. *J. Forensic Account.* **5**(1), 17–34 (2004)
13. Haroush, M., Frostig, T., Heller, R., Soudry, D.: A statistical framework for efficient out of distribution detection in deep neural networks. In: International Conference on Learning Representations (2021)
14. Hashemi, V., Křetínský, J., Mohr, S., Seferis, E.: Gaussian-based runtime detection of out-of-distribution inputs for neural networks. In: Feng, L., Fisman, D. (eds.) RV 2021. LNCS, vol. 12974, pp. 254–264. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-88494-9_14
15. Hendrycks, D., Dietterich, T.: Benchmarking neural network robustness to common corruptions and perturbations. In: International Conference on Learning Representations (2018)
16. Hendrycks, D., Gimpel, K.: A baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv preprint [arXiv:1610.02136](https://arxiv.org/abs/1610.02136) (2016)
17. Hendrycks, D., Mazeika, M., Dietterich, T.: Deep anomaly detection with outlier exposure. arXiv preprint [arXiv:1812.04606](https://arxiv.org/abs/1812.04606) (2018)
18. Henzinger, T.A., Lukina, A., Schilling, C.: Outside the box: abstraction-based monitoring of neural networks. arXiv preprint [arXiv:1911.09032](https://arxiv.org/abs/1911.09032) (2019)
19. Huang, R., Geng, A., Li, Y.: On the importance of gradients for detecting distributional shifts in the wild. *Adv. Neural. Inf. Process. Syst.* **34**, 677–689 (2021)
20. Koh, P.W., et al.: Wilds: a benchmark of in-the-wild distribution shifts. In: International Conference on Machine Learning, pp. 5637–5664. PMLR (2021)

21. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**(6), 84–90 (2017)
22. Lam, E.Y., Goodman, J.W.: A mathematical analysis of the DCT coefficient distributions for images. *IEEE Trans. Image Process.* **9**(10), 1661–1666 (2000)
23. Lin, J.: Divergence measures based on the shannon entropy. *IEEE Trans. Inf. Theory* **37**(1), 145–151 (1991)
24. Milani, S., Tagliasacchi, M., Tubaro, S.: Discriminating multiple JPEG compressions using first digit features. *APSIPA Trans. Sig. Inf. Process.* **3**, e19 (2014)
25. Nguyen, A., Yosinski, J., Clune, J.: Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 427–436 (2015)
26. Olber, B., Radlak, K., Popowicz, A., Szczepankiewicz, M., Chachula, K.: Detection of out-of-distribution samples using binary neuron activation patterns. *arXiv preprint [arXiv:2212.14268](https://arxiv.org/abs/2212.14268)* (2022)
27. Pasquini, C., Boato, G., Pérez-González, F.: Multiple jpeg compression detection by means of benford-fourier coefficients. In: *2014 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 113–118. IEEE (2014)
28. Pérez-González, F., Heileman, G.L., Abdallah, C.T.: Benford’s law in image processing. In: *2007 IEEE International Conference on Image Processing*, vol. 1, pp. I–405. IEEE (2007)
29. Raimi, R.A.: The first digit problem. *Am. Math. Mon.* **83**(7), 521–538 (1976)
30. Rao, K.R., Yip, P.: *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Academic Press, San Diego (2014)
31. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**(3), 211–252 (2015). <https://doi.org/10.1007/s11263-015-0816-y>
32. Sahu, S.K., Java, A., Shaikh, A., Kilcher, Y.: Rethinking neural networks with benford’s law. *arXiv preprint [arXiv:2102.03313](https://arxiv.org/abs/2102.03313)* (2021)
33. Schirrmeister, R., Zhou, Y., Ball, T., Zhang, D.: Understanding anomaly detection with deep invertible networks through hierarchies of distributions and features. *Adv. Neural. Inf. Process. Syst.* **33**, 21038–21049 (2020)
34. Singh, N., Bansal, R.: Analysis of benford’s law in digital image forensics. In: *2015 International Conference on Signal Processing and Communication (ICSC)*, pp. 413–418. IEEE (2015)
35. Stano, M., Benesova, W., Martak, L.S.: Explaining predictions of deep neural classifier via activation analysis. *arXiv preprint [arXiv:2012.02248](https://arxiv.org/abs/2012.02248)* (2020)
36. Steinhart, J., Charikar, M., Valiant, G.: Resilience: a criterion for learning in the presence of arbitrary outliers. *arXiv preprint [arXiv:1703.04940](https://arxiv.org/abs/1703.04940)* (2017)
37. Yang, J., Zhou, K., Li, Y., Liu, Z.: Generalized out-of-distribution detection: a survey. *arXiv preprint [arXiv:2110.11334](https://arxiv.org/abs/2110.11334)* (2021)
38. Zhang, M., Zhang, A., Xiao, T.Z., Sun, Y., McDonagh, S.: Out-of-distribution detection with class ratio estimation. *arXiv preprint [arXiv:2206.03955](https://arxiv.org/abs/2206.03955)* (2022)



Autonomous Vehicle Algorithm Decision-Making Considering Other Road Users

Ishan Tyagi and Subramaniam Ganesan^(✉)

Oakland University, Rochester Hills, MI 48307, USA
{ityagi, ganesan}@oakland.edu

Abstract. A new method is proposed to assess the threat level for the traffic scenarios that have a large lateral offset target or host significant turning, with the consideration of the predicted host heading angle change, host vehicle dimensions information, and host velocity reference point migration. Based on that, this system creates a new predicted longitudinal and lateral offset calculation and thus proposed a new target in the path check algorithm to detect when the target will be present in the host-predicted path based on the improved path prediction method. This method also provides a target in the path check algorithm that determines the threat potential and threat level based on the predicted lateral offset and longitudinal time to collision (TTC long). This function can provide the system to work for stationary targets and also prevent false brakes for a target that could generate a small near miss situation, e.g. a pedestrian could generate a 1 m near miss.

Keywords: Autonomous vehicle · Path check algorithm · Assess the threat level

1 Introduction

Operating a motor vehicle and sharing the roadways with other road users entails a certain amount of risk. The legal system helps to manage the risk by placing a duty of care on each driver and clarifying the expected behavior for tasks such as following lane changes or navigating at intersections in specific provisions of the traffic code [1].

Active safety functions in general need a decision module, i.e., the function needs to decide if and when to intervene or warn the driver. This is typically done by computing a threat level based on the available information and then comparing this to a threshold. The threat level computing is based on the internal vehicle signals, but in many currently emerging and future active safety systems, the threat level is based on the position of surrounding vehicles relative to the host vehicle and relative to the lane [2].

Considered a simple road traffic scenario where the autonomous vehicle algorithm will challenge the basic traffic rules for the intersections. While considering the scenarios there are happy paths and critical paths which will challenge the algorithm for the basic traffic rules consideration. The basic scenario for the autonomous vehicle is to be considered there will be object classification, object detection, and collision avoidance.

Collisions at intersections account for 41% of all police-reported crashes, 46% of all injuries and 21% of all fatalities [3]. The simplified motion model makes it possible to

derive an analytical solution. In this paper, I am assisting human drivers with negotiating busy intersections in the presence of possible errant drives with uncertain intentions [4].

2 Outline

The key idea in this paper is to assess the collision threat of surrounding vehicles in real-time at each sampling time instance to evaluate the threat level for any detected target in the sensing field of view of the host vehicle (Fig. 1).



Fig. 1. Target vehicle/pedestrian entering host path

This is not specific to the intersection but to any given scenario of the trajectory of vehicles. In Section 2, we describe decision-making when to apply autonomous braking. Section 3 contains the proposed method for estimating for the driver to brake or accelerate. Section 4 describes the braking system method and simulations. Section 5 presents Conclusion.

3 Research on New Recommendation Algorithm

The proposed method for the algorithm is described on the bases of the flow chart enplaning all the required steps for the algorithm to perform when the target vehicle of the pedestrian is in the path of the host vehicle (Fig. 2).

The Target sensing system will down-select the target first on the long-range down selection and second for the near-range down selection (NRDS). After that, the target & host relative position prediction is done. Now it's to calculate the Time to collision (TTC). Now have to calculate the predicted offsets and safety margins at the TCC time point. With this also will check for the collision potential determination & Target in

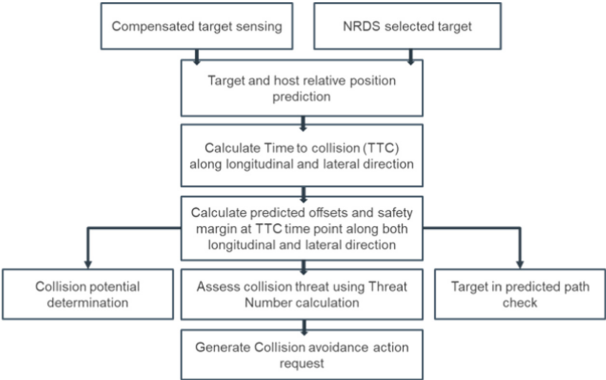


Fig. 2. Overall flow chart for the threat assessment algorithm.

the predicted pat check. After all the consideration and calculation and considering the traffic code collision avoidance action request is given first to the driver, if a driver doesn't interfere then the autonomous vehicle brakes accordingly [5–7].

3.1 Threat Assessment

Considering the sensing for pedestrians and target vehicles on the assumption of the sensor set for the autonomous vehicle.

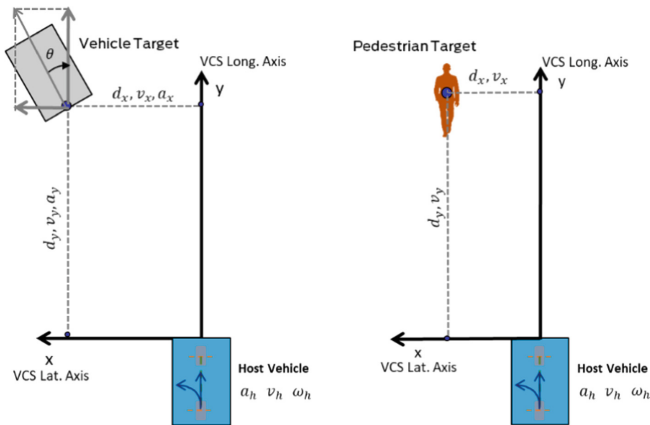


Fig. 3. Sensing system interface

The notations in Fig. 3 are given below.
 v_h : host vehicle longitudinal velocity (m/s) at the current time.
 a_h : host vehicle longitudinal acceleration (m/s²) at the current time.
 ω_h : host vehicle yaw rate (rad/s) at the current time.

d_x : target relative lateral position in the host vehicle coordinate system (VCS) (m) at the current time.

v_x : target relative lateral velocity in the host VCS (m/s) at the current time.

a_x : target relative lateral acceleration in the host VCS (m/s²) at the current time.

d_y : target relative longitudinal position in the host VCS (m) at the current time.

v_y : target relative longitudinal velocity in the host VCS (m/s) at the current time.

a_y : target relative longitudinal acceleration in the host VCS (m/s²) at the current time.

θ : target relative body heading angle in the host VCS (rad) at the current time.

In terms of target classification, the sensing signals use a different interface to provide robust functionality. For example, the pedestrian target doesn't use target acceleration and heading angle information but which are available for a vehicle target. In order to assess the threat level for both target type, the new method considers all available sensing signals.

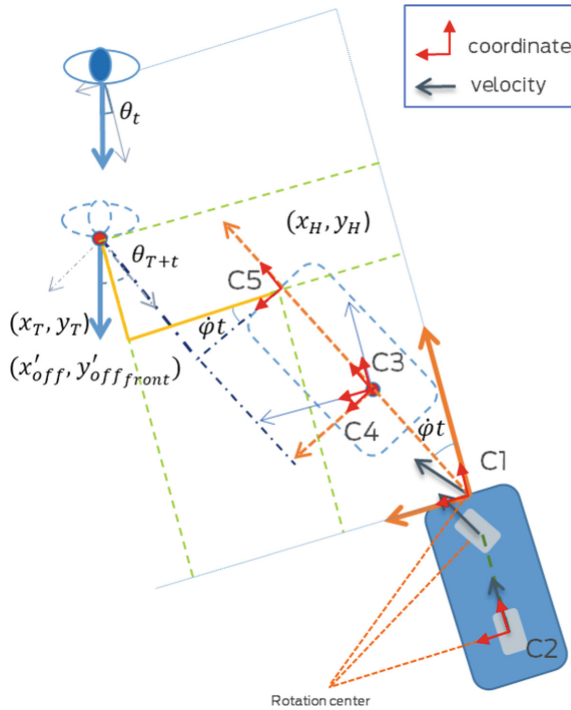


Fig. 4. Longitudinal & lateral offset prediction

Figure 4 has a total 5 different coordinates C1, C2, C3, C4, and C5 are used to predict the longitudinal and lateral offset for future time t and current time T . The 5 coordinates are defined as follows.

C1: Host vehicle ISO coordinate system (VCS) referred to host front bumper center at current time T .

C2: Host vehicle VCS referred to host rear wheel axle center at current time T .

C3: Host vehicle VCS referred to host rear wheel axle center at future time $(T + t)$ without transform into predicted host heading angle.

C4: Host vehicle VCS referred to host rear wheel axle center at future time $(T + t)$ with transform into predicted host heading angle.

C5: Host vehicle predicted VCS referred to host front bumper center at future time $(T + t)$.

Host Path Prediction:

For the host path prediction, a constant yaw rate and acceleration (CTRA) motion model is used for host path prediction in the C2 coordinate.

To have the accurate curvature estimation for the target vehicle, a constant acceleration model is adopted for target vehicle. The predicted lateral and longitudinal distance between host & Target vehicle at time $t + T$ is written in the equation [8–11]

$$L_{LatRel}(t + T) = -\frac{1}{24}v_{LongHost}(t)\dot{\psi}_{Host}^3(t) * T^4 + \frac{1}{3}a_{LongHost}(t)\dot{\psi}_{Host}(t) * T^3 \\ + \frac{1}{2}(v_{LongHost}(t)\dot{\psi}_{Host}(t) + a_{LatTrgt}(t) - a_{LatHost}(t)) * T^2 \\ + (v_{LatTrgt}(t) - v_{LatHost}(t)) * T + L_{LatRel}(t)$$

Therefore, at time t , the host longitudinal and lateral prediction position is shown as (1), (2):

$$y_H(t) = \int \cos(\omega_h t)(a_H t + v_{H,0})dt = \frac{\omega_h \sin(\omega_h t)(a_H t + v_{H,0}) + a_H \cos(\omega_h t)}{\omega_h^2} \quad (1)$$

$$x_H(t) = \int \sin(\omega_h t)(a_H t + v_{H,0})dt = \frac{a_H \sin(\omega_h t) - \dot{\varphi} \cos(\varphi \dot{\omega}_h t)(a_H t + v_{H,0})}{\omega_h^2} \quad (2)$$

The notations used in the (1), (2) are given below:

$y_H(t)$: Host vehicle predicted longitudinal position at time t in C2 coordinate.

$x_H(t)$: Host vehicle predicted lateral position at time t in C2 coordinate.

Target Prediction:

Since target sensing signals are refer to C1 - host front bumper center in VCS and host kinematic signal is referred to host rear axle center, to ensure the host prediction and target prediction are calculated in the same coordinate system, the target sensing signals need to translate to rear axle center coordinate C2. Therefore, the target kinematic position signals in C2 coordinate are shown as (3), (4):

$$d_y^{C2}(t) = d_y(t) - L \quad (3)$$

$$d_x^{C2}(t) = d_x(t) \quad (4)$$

The notations used in the (3), (4) are given below:

t : Current time instance.

d_y^{C2} : Target relative longitudinal position in C2 coordinate.

d_x^{C2} : Target relative lateral position in C2 coordinate.

L : Distance from host front bumper center to host rear wheel axle center.

For target prediction, a constant acceleration (CA) kinematic motion model is used for target path prediction in the C2 coordinate. Therefore, the target longitudinal and lateral prediction is shown as before as sensing interface definition

$$y_T(t) = \frac{1}{2}a_y t^2 + v_y t + d_y^{C2} \quad (5)$$

$$x_T(t) = \frac{1}{2}a_x t^2 + v_x t + d_x^{C2} \quad (6)$$

The notations used in (5), (6) are given below:

$y_T(t)$: Target vehicle predicted longitudinal position at time t in C2 coordinate.

$x_T(t)$: Target vehicle predicted lateral position at time t in C2 coordinate.

Final Predicted offset:

Eventually, translate the coordinate from C4 to C5 to be aligned with the sensing system interface definition at future time t and the finalized predicted longitudinal and lateral offset at time t will be expressed as (7), (8):

$$y_{offset}(t) = y_{offset}(t)^{c4} - L \quad (7)$$

$$x_{offset}(t) = x_{offset}(t)^{c4} \quad (8)$$

Combined (1)–(8), the relative longitudinal and lateral offset prediction can be expressed as:

$$y_{offset}(t) = \begin{bmatrix} \frac{a_x \sin(\omega_h t) + a_y \cos(\omega_h t)}{2} \\ \cos(\omega_h t)v_y + \sin(\omega_h t)v_x \\ \frac{a_H \cos(\omega_h t) - a_H - v_H \omega_h \cdot \sin(\omega_h t) + d_y \omega_h^2 \cdot \cos(\omega_h t) + d_x \omega_h^2 \cdot \sin(\omega_h t)}{\omega_h^2} \end{bmatrix} \cdot [t^2 t^1 t^0] \quad (9)$$

$$x_{offset}(t) = \begin{bmatrix} \frac{a_x \cos(\omega_h t) - a_y \sin(\omega_h t)}{2} \\ \frac{a_H}{\phi} + \cos(\omega_h t)v_x - \sin(\omega_h t)v_y \\ \frac{a_H \sin(\omega_h t) - v_H \omega_h + v_H \omega_h \cdot \cos(\omega_h t) - d_x \omega_h^2 \cdot \cos(\omega_h t) + d_y \omega_h^2 \cdot \sin(\omega_h t)}{\omega_h^2} \end{bmatrix} \cdot [t^2 t^1 t^0] \quad (10)$$

TTC and Predicted offset:

Define longitudinal TTC as the time when the host and target reach the same longitudinal position and lateral TTC as the time when the host and target reach the same lateral position. Thus the TTC along longitudinal and lateral direction can be defined as (11), (12):

$$y_{offset}(TTC_{long}) = 0 \quad (11)$$

$$x_{offset}(TTC_{lat}) = 0 \quad (12)$$

To acquire the TTC along both directions, The $y_{offset}(t)$ and $x_{offset}(t)$ are required to be solved efficiently in real time, thus the Taylor series expansion is being used to substitute the trigonometry terms in (11), (12). Based on the designated traffic scenarios, 2 conditions need to be met to determine the order of Taylor series expansion.

1. The accurate approximation range of Taylor series expansion should cover the usage range of host prediction heading angle and time to collision for designated traffic scenarios
2. Use minimal order for approximation to simplify the calculation in real time.

Therefore, the order of Taylor series expansion is determined as following expression for this application [12–14].

$$\sin(\omega_h t) = \omega_h t - \frac{(\omega_h t)^3}{6} \quad (13)$$

$$\cos(\omega_h t) = 1 - \frac{(\omega_h t)^2}{2} + \frac{(\omega_h t)^4}{24} \quad (14)$$

Threat Number calculation:

From the Threat Assessment for avoiding collisions with perpendicular vehicles at Intersections paper the calculation of the BTN (Brake Threat Number).

The host BTN will be calculated as:

$$BTN_{host} = \min\left(\frac{v_t(t)}{TTC_{long}(t) - T_{delay}(t)} * \frac{1}{decel_{host}^{max}}, 1\right) \quad (15)$$

Also a new position based BTN is proposed to calculate the needed deceleration to stop the host vehicle at a certain safety distance along the longitudinal direction ahead of collision to improve the robustness of system, the calculation is shown as (16):

$$BTN_{dis} = \min\left(\frac{2 * \max(LongSafetyMarg(t) - |L_{predLat}(t)|, 0)}{(TTC_{long}(t) - T_{delay}(v_h))^2} * \frac{1}{decel_{host}^{max}}, 1\right) \quad (16)$$

Within (23), $LongSafetyMarg(t)$ is a tunable parameter in terms of longitudinal TTC depends on the requirement of system. $T_{delay}(v_h)$ is also a tunable parameter in terms of host vehicle velocity and specific brake system.

Eventually, the final BTN will be $BTN = \min(BTN_{host}, BTN_{dis})$.

Target in path check:

When $TTC_{long}(t) < TTC_{thresh}$:

A function based on predicted lateral distance is defined to represent the possibility that target will be present in host path:

$$f_{statPed}(L_{predLat}(t)) = e^{\frac{-(L_{predLat}(t))^2}{2\sigma^2}} \quad (17)$$

In (15):

$$\sigma = 1.154 * \left(\frac{Hostwidth}{2} + safe_{margin}(t) \right) \quad (18)$$

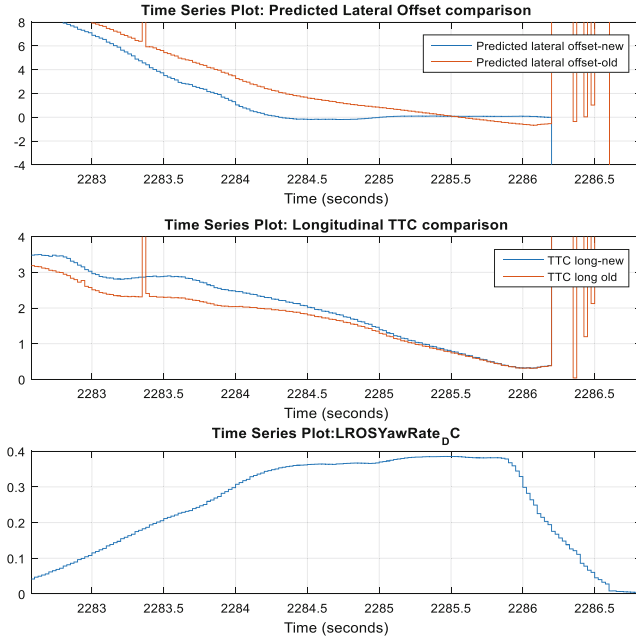


Fig. 5. 50% collision test pedestrian target

4 Experimental Results and Analysis

The collision avoidance system has been evaluated on a traffic scenario with the Matlab Model and Simulink for the simulation run.

The test is considered with the vehicle being impacted at the 50% collision (Fig. 5). 1 m near miss pass scenario pedestrian target (Fig. 6).

For the stationary Target test, the predicted offset shows the same consistency as the moving target (Fig. 7).

From all the above calculations and the algorithms the Threat number is identified for the vehicle to miss the target or automated braking is applied to avoid the collision.

If the offset is incorporated in the threat calculation the difference in the impact or safe can be seen (Figs. 8 and 9).

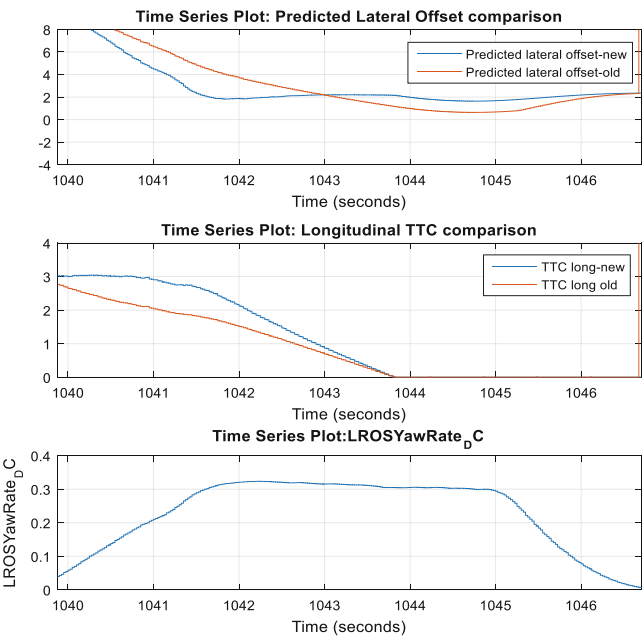


Fig. 6. 1 m miss pedestrian target

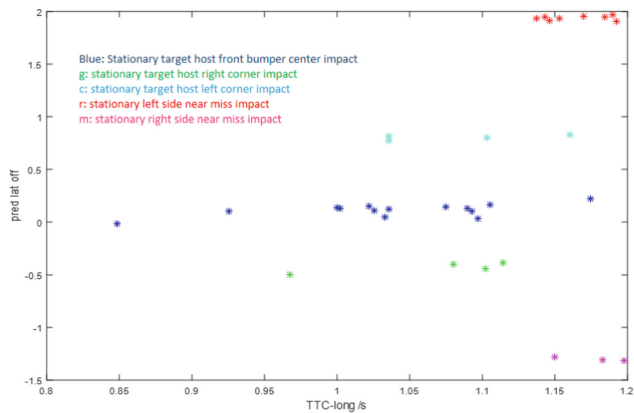


Fig. 7. Stationary Target predicted offset comparison based on the collision point

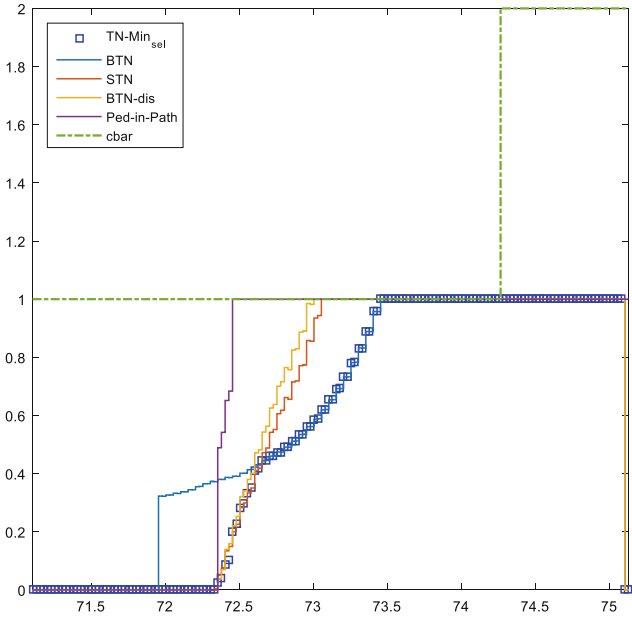


Fig. 8. Threat number when impact

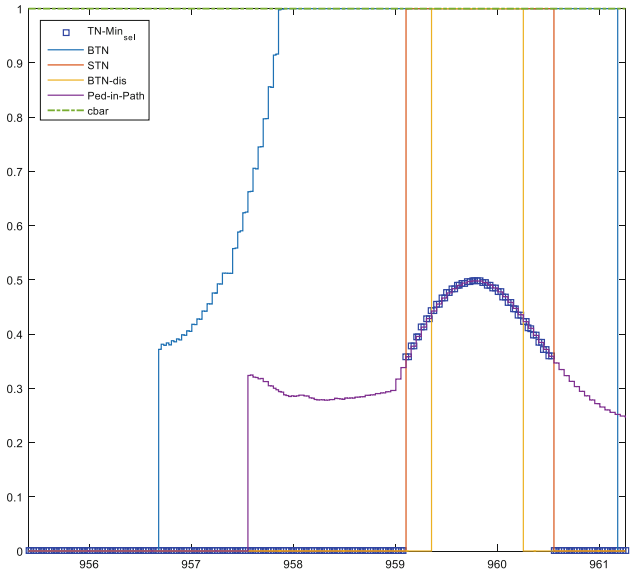


Fig. 9. Threat number when 1m near miss offset

5 Conclusion

In conclusion, Predicted lateral offset is getting accurate much faster compared with the previous predicted lateral offset calculation method, especially at the initial detection phase where larger lateral offset and bigger vehicle heading angle change when collision happen. The new approach will detect the collision potential much earlier than previous method by comparing the predicted lateral offset at longitudinal TTC time instance with safety margin threshold. More distinguishable between impact scenarios and near miss scenarios due to the consistency improvement of the predicted lateral offset. Capable to detect a potential collision target as much as 1.5 s ahead compared with previous method, which could provide the potential capability to provide warning or other prevention actions earlier without generate the unexpected false alarm. The result of longitudinal TTC and predicted lateral offset shows improvement of the purposed method in terms of the host and target prediction perspective. The result of final threat number shows how the improved prediction and the target in path check algorithm help to improve system performance.

References

1. IEEE P2846 working group, Draft Standard for Assumptions for Models in Safety-Related Automated Vehicle Behavior, p. 2846. IEEE (2021)
2. Bonnefon, J.F.: The social dilemma of autonomous vehicles. *Science* **352**, 1573–1576 (2016)
3. Tyagi, I.: Threat assessment for avoiding collisions with perpendicular vehicles at intersections. In: 2021 IEEE International Conference on Electro Information Technology (EIT), pp. 184–187 (2021)
4. Dahl, J., de Campos, G.R., Olsson, C., Fredriksson, J.: Collision avoidance: a literature review on threat-assessment techniques. *IEEE Trans. Intell. Veh.* **4**(1), 101–113 (2019)
5. Aoude, G.S., Luders, B.D., Lee, K.K.H., Levine, D.S., How, J.P.: Threat assessment design for driver assistance system at intersections. In: 13th International IEEE Conference on Intelligent Transportation Systems, pp. 1855–1862 (2010)
6. Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida (2016)
7. Tracking and Threat Assessment For Automotive Collision Avoidance. Andreas Eidehall
8. Schubert, R., Richter E., Wanielik, G.: Comparison and evaluation of advanced motion models for vehicle tracking. In: Proceedings of the 11th International Conference on Information Fusion (2008)
9. Li, X.R., Jilkov, V.P.: Survey of maneuvering target tracking. part i: dynamic models. *IEEE Trans. Aerosp. Electron. Syst.* **39**(4) (2003)
10. Gustafsson, F.: Statistical Sensor Fusion. Studentlitteratur (2018)
11. Svensson, D.: Derivation of the discrete-time constant turn rate and acceleration motion model. In: 2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF), pp. 1–5 (2019)
12. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Section 5.6 Quadratic and Cubic Equations, Numerical Recipes: The Art of Scientific Computing, 3rd edn. Cambridge University Press, New York. ISBN 978-0-521-88068-8
13. Yoon, Y., Kim, C., Lee, J., Yi, K.: Interaction-aware probabilistic trajectory prediction of cut-in vehicles using gaussian process for proactive control of autonomous vehicles. *IEEE Access* **9**, 63440–63455 (2021)
14. Toghi, B.: A maneuver-based urban driving dataset and model for cooperative vehicle applications. In: 2020 IEEE 3rd Connected and Automated Vehicles Symposium (CAVS), pp. 1–6 (2020)



The Role of Administrative Leadership in Crisis Management in the Jordanian Ministry of Planning and International Cooperation

Haron Ismail Al-lawama¹, Saddam Rateb Darawsheh² (✉) ,
Mohammad Salameh Zaid Almahairah³, Anwar Saud Al-Shaar⁴,
Asmaa Jumaha AlMahdawi⁵ , Azhar Shater⁵ , Imen Gmach⁶,
Hamed Omar Abdalla⁷, and Muhammad Turki Alshurideh⁸

¹ Universiti Malaysia Terengganu, Kuala Terengganu, Terengganu, Malaysia

² Department of Administrative Sciences, The Applied College, Imam Abdurrahman Bin Faisal University, P.O. Box: 1982, Dammam, Saudi Arabia
srdarawsehe@iau.edu.sa

³ Department: Management Information System, Asra University, Amman, Jordan
mohammad.almahirah@iu.edu.jo

⁴ Imam Abdurrahman Bin Faisal University, Dammam, Saudi Arabia

⁵ Department of Humanities and Social Sciences, College of Education, Al-Ain University, Al-Ain, United Arab Emirates
{Asmaa.almahdawi, Azhar.shater}@aau.ac.ae

⁶ Administrative Sciences Department, Imam Abdurrahman Bin Faisal University, Dammam, Saudi Arabia
Ihgmach@iau.edu.sa

⁷ Collaborator with Jouf University/College of Business Administration, Sakakah, Saudi Arabia

⁸ Department of Marketing, School of Business, The University of Jordan, Amman 11942, Jordan
m.alshurideh@ju.edu.jo

Abstract. The main objective of the current study is to find out the relationship between leadership style and crisis management in the Jordanian Ministry of Planning and International Cooperation for the department head, managers, and senior management in the Jordanian Ministry of Planning. A quantitative methodology was used to track the goals, in which people who participated in the survey, received the highest score ever Leadership type, the study concluded that transformational leadership has become more stringent in recent decades than transactional leadership, but democratic leadership has no significant relationship with crisis management, and often the growth strategies of those organizations need to change in times of crisis, and the results from the current research will serve as a guide for the authorities government.

Keywords: Transformational leadership · Crisis management · Jordan ministry of planning and international cooperation

1 Introduction

The leader's talent lies in the leader's ability to manage crises, the human component ranks first among various other production factors and contributes to the achievement of the facility's intended goals (Albarode 2015; Ahmed et al. 2021) cited from Gardner Willawan (2014), linearity based on ingenuity and creativity in the relationship between the financial situation and society, leadership to influence common goals is the process of persuasion, the way an individual (or leadership team) persuades a group to pursue a specific goal, current financial situation (Al Shraah et al. 2022), leadership refers to business leaders, large organizations who regularly make important decisions and expect positive results in achieving their goals (Alameeri et al. 2020; Odeh et al. 2021), I strongly recommend it, Goal achievement for government officials and administrators (Akoi et al. 2021), Decision-making takes place to give orders, solve problems and manage crises (Ali & Anwar 2021), deal with a large number of unsatisfactory results, and the most important thing is to make the strongest decisions that all owners of the organization fully understand (Ali et al. 2021; Alshurideh et al. 2022), this raises questions and the focus of this study is leadership in crisis management, style and leadership, it has a significant impact on shaping organizational culture (Andovar et al. 2020). In general, organizational culture creates value for work, and encourages good behavior among.

Employees, which are expressed by leaders (Abuhashesh et al. 2021; Al-bawaia et al. 2022), in today's world, organizations try to group their people as a first step towards overcoming crises in general that may harm them in the future, many previous scholars have studied Leadership and distinguished well between leadership and managerial thinking with regard to achieving results and organizational effectiveness (Anwar and Gahor 2017; AlShehhi et al. 2020), different scholars and researchers have used perceptions and theories about organizational behavior.

Organizational behavior consists of leadership and employee styles and depends on the personality of the leader and the ability to deal with problems to accommodate diversity, economic strength, enhance professional skills, create work-life balance, and improve customer service, encourage ethical behavior, provide service, crises and deal with them in a scientific way (Agha et al. 2021; Sabir et al. 2021; Alshurideh 2022). Leadership refers to business leaders, large organizations who regularly make important decisions and expect positive results in achieving their goals, I strongly recommend it, goal achievement for government officials and administrators (Akoi et al. 2021), decisions are constantly being made to give orders, solve problems, manage crises (Ali & Anwar 2021), and deal with multiple unsatisfactory outcomes, the most important thing is to make the best decisions that are fully understood by all stakeholders (Ali et al. 2021).

This raises questions and the focus of this study is leadership in crisis management, style and leadership, it has a significant impact on shaping organizational culture (Andavar et al. 2020), Organizational culture creates value, encourages good behavior among members, and is articulated by leaders, organizations are looking to assemble their teams as one step toward overcoming crises that could hurt them, as many previous scholars have, Studied leadership and have made a good distinction between leadership and managerial thinking in relation to achieving results (Sabir et al. 2021).

2 Literature Review

Leadership, according to Anwar (2017), is defined as the ability of a person to influence the rest of a group of individuals in order to achieve a specific goal or set of goals. Leadership is defined through the following four actions: the ability to adapt and the ability to change, commitment to the company and organizational vision, the ability to determine and know the direction, and the development of the company's mission. For work, this aspect shows strength in individual or group operations, and can be important if the organization tries to develop itself.

To identify individual and group competencies, to provide rewards or as a measure for the successful completion of certain task, according to Anwar and Balcioglu (2016), Transactional leaders are directed to individuals by describing their responsibilities to achieve organizational goals, according to transactional leadership, they help individuals identify and assign duties to each employee in order to achieve the company's goals and increase its productivity (Anwar and Shukur 2015), which affects transformational leadership and functions, improving transactional leadership methods for general positions, namely Task Mostly in subcategories, people do passive action to manage transactions of adjustment This variation is called living form of transactional leadership styles (Avolio and Bass 2004), Leaders who encourage their followers to prioritize the interests of the organization over individual interests, and leaders who care about the interests of their followers are considered a good result, the researchers considered transformational leadership from many approaches in which transformational leadership is manifested when a crew member performs excessive work, exercises, and discloses his mission, function, and purpose, and is noted by the desire of each member to prioritize the interest of the organization over personal interests (Bass 1990), the crisis. It is a tense situation that threatens the higher goals of the institutions, but rather endangers them and threatens their survival, and requires immediate action. Confidence and time pressure, covering a limited period of time to remove or reduce the effects through the necessary precautions, a crisis can occur as a result of an unpredictable event or an unexpected outcome of an event that has been considered a risk, and crises almost always require that decisions be taken quickly to limit the damage to organization (Anwar & Climis 2017).

2.1 Transformational Leadership

The increasing trend of the central role of knowledge and economic knowledge as the central lever of organizations in a competitive environment has created new demands and a new atmosphere in the field of organizational activity (Al Kurdi et al. 2020; Alshurideh et al. 2022a, b), which has inspired the development of the theory of transformational leadership, rice field, there are transformations and renovations in the organization, it is the kind of guidance that allows hidden assets to expand their potential and move into action, organizations cannot form such strategic coalitions without emphasizing transformative leadership styles (Braun et al. 2006; Al-Dhuhouri et al. 2020), kanger and canangoargue that transformative leaders express perspectives that generate new ideas and opportunities for the future in times of crisis and change, and he often generates new strategic dimensions barnes showed that transformative leaders have cognitive powers

and persuade others to ask them to perform unusual tasks. Barnes argues only transformational leaders can chart the essential path of the new organization, because they are the source of change, they fully understand organizational change and are fully capable of directing waves of change a transformational leader articulates the organization's future vision he presents a model consistent with that vision he embraces the group's goals and supports each individual (Berson and Avolio 2004).

Style (Bass 1990), transformational leadership styles like this are meant to instill a long-term questioning mentality in leaders, they have long-term goals in mind that encourage employees to do more than they used to, reshaping and expelling vendors instead of simply tracking and managing them this is how transformational leadership occurs most easily, when leaders and employees support each other, elevate each other to higher levels of ethics, values and motivation, and then deliver the highest levels of performance (Burns 1978), in this respect, in agreement with Burns (1978, pp. 6–11), the end result of transforming leadership is to transform employees into leaders, it is the courtship of mutual stimulation and promotion that can transform a leader into an ethical salesperson, the reason for this is that transformation affects all managers and employees transformational leadership is a key outcome of leaders who inspire followers to prioritize group interests over individual interests.

2.2 Democratic Leadership

A leader who believes that others can participate in decision-making and hold them accountable, this is the premise for building the engagement of others and the beliefs of those around them, whether they exercise leadership through consensus or consensus, can be chosen on the premise that it may contradict laws within an organization, empower society by changing the landscape of the law and expanding it according to its guidelines, she is inspired to work hard to make her own dreams come true, she was promised this type of manager values responsibility, values collective participation, likes to study others, and often has great and innovative bosses who are hired to make their dreams come true (Hussain A. Sinjar Alsamary 2014).

Democratic leaders can take responsibility for their mistakes and accept the consequences, unlike autocratic bosses they are often above the law does leadership that believes in competence have the right of others to participate in and be held accountable for the decision-making process it must represent the goals to be achieved, this for the basis, by choice or agreement for others to build around them the participation and beliefs to take leadership positions.

Programs and projects may be chosen on a democratic basis and that many of the laws followed in the organization or society and that they must comply with the laws of the organization and to change the rules followed and formulate policies to achieve the promised goals (Al-Samarai 2014) affects his followers using a democratic method to achieve amazing results for democratic leaders and to achieve future success for employees and for the organization, leaders are willing to take any risks to prove their commitment, according to previous research (Anwar and Qadir 2017) and others (Anwar 2017) Leadership refers to an individual's ability to influence other groups of people to achieve a goal or set of goals, and thus, the Correlations between charismatic leadership and one's OCB, (Anwar & Balcioglu 2016) provided findings from other studies confirming this

Their programs and projects selected on the basis that they violate several prevailing in the organization and they must comply with to change and formulate policies to achieve the promised goals (Alsamaray 2014) influences his followers achieve amazing results democratic leaders have achieve future success of their supporters, moreover, leaders are willing to take any demonstrate their commitment, according to previous research (Anwar and Qadir 2017) and others (Anwar 2017) leadership refers to an individual's ability to influence other groups of people to achieve a goal or set of goals consequently, leadership is all about motivating individuals to replicate change and vision (Anwar & Louis 2017), we found a positive association between democratic leadership and OCB Further (Anwar and Balcioglu 2016), found a relationship between charismatic leadership and organizational success Furthermore, Anwar & Ghafoor (2017) and Anwar & Climis (2017), found he positive associations between charismatic leadership and his OCB, additionally (Anwar & Balcioglu 2016) presented results from other studies.

2.3 Transactional Leadership

Transaction leadership is part of a leadership style that focuses on oversight, organization, and performance; it is an integral part of the overall leadership model, this type of management emerged as a source of competitive advantage during the Industrial revolution; typical tactics for this type of management are strategy, efficiency goals, economies of scale, and quality differentiation, transactional managers focus on performance-related tasks and goals (Stajkovic, Alex 2019).

Transactional leadership is a type of leadership style in which submissive leaders encourage compliance through rewards and punishments. Through a system of rewards and punishments, leaders can generally motivate their followers in the short term and long term. In contrast, transformational leaders display behaviors associated with constructive and corrective transactions, the style is named Reward construction and correction method for the organization is labeled Management-by-Exception (active and passive), transactional leadership defines expectations and promotes performance to achieve these levels, contingent reward and management exception are two core behaviors associated with 'management' functions in organizations, full range leaders do this and more (Bass & Avolio 2004), transformational leadership acts as a bridge between leaders and followers to develop clear understanding of follower's interests, values and motivational level, It helps leaders achieve their goals by working within the organizational framework, and encourages leaders to be expressive and adaptive to new and improved practices and changes in the environment, where the method of work is organized and divided into principles called positive transactional leadership, collective transactional leadership, and the same function, which achieve distinctive leadership goals in environments Varied work, for example, the proactive form of transactional leadership style is best for management functions such as defining Bass (2004).

2.4 Crisis Management

A crisis is defined as a state of tension to which an organization or individuals are exposed within the organization, which causes the higher goals in the public sectors to be endangered. Rather, it endangers the existence and survival of the organization, and therefore requires the organization to take quick and immediate measures, in addition to a situation in which preventive measures are not sufficient, or predict the future (Anwar & Climis 2017). The crisis management team is considered the highest authority in the organization, which deals with any threat to the organization and is able to implement immediate and rapid actions under pressure in the shortest possible time. This method of crisis management is one of the most widespread and legitimate. According to the classification of all the adverse conditions observed within the organization as a crisis is not the right strategy, for example, the governance that assumes that the rate of inflation next year will be much lower than the rate of this year may encounter problems and classify the situation as a crisis, The crisis must be viewed as a leadership issue. A financial challenge and loss that does not threaten the existence of the organization or an imbalance in its position cannot be considered a crisis. It is wrong to mention general leadership issues that lead to low morale and low motivation (which cause problems for the company and its managers), the ability to collect accurate and comprehensive information, and information about the problems that the organization may face) (Anwar & Qadir 2017).

The crisis is preceded by warning signs, but it can also occur as a result of a series of unexpected events, and in this regard, crises can be divided into types of re-crisis stage, and this first stage occurs before the crisis itself reaches the acute crisis stage. Stage This is the point where your organization begins to notice the first signs of an unfolding crisis Response stage Chronic stage (Anwar & Surarchith 2015) A crisis can be caused by structured conditions, in other words, such as rigid organizational structures, communication disorders and managerial shortcomings are among the potential crises; The main reason for the occurrence of leadership crises is deficiencies and errors in the leadership style, and therefore many internal problems in the leader's style can be identified as reasons that can lead to failure and crises (Anwar 2016).

3 Research Hypotheses

3.1 Theoretical Framework and Hypotheses Development

H1: There is a significant effect of transformational leadership on crisis management of the Jordanian ministry of planning and international cooperation.

H2: There is a significant effect of transactional leadership on crisis management of the Jordanian ministry of planning and international cooperation.

H3: There is a significant effect of democratic leadership on crisis management of the Jordanian ministry of planning and international cooperation (Fig. 1).

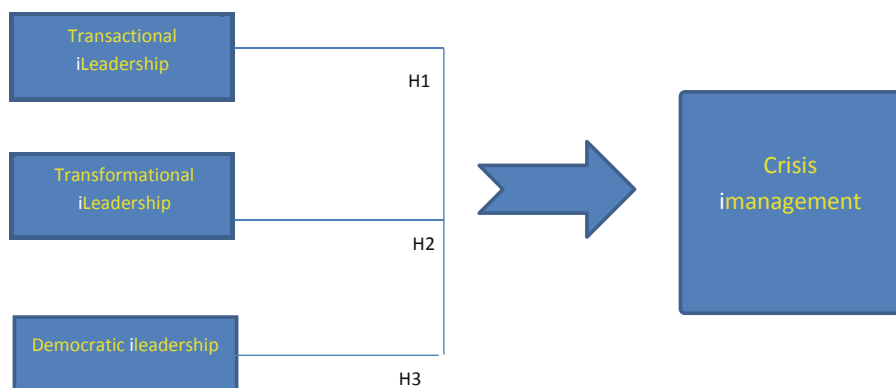


Fig. 1. Conceptual frame work of the research

4 Methodology

Study Design

This research was examined by the researchers using quantitative methods, and a questionnaire was made to collect the necessary data for the research, according to (Kumar 2005), and therefore this research can be described as a cross-sectional study based on the number of potential contacts from the selected study population.

4.1 Survey Methodology

A number of department heads, directors, and senior management in the Jordanian ministry of planning associated with this study, and the Krejcie and Morgan table, were used to determine the appropriate sample size for this study, according to Krejcie and Morgan (1970), for a population of 67, the appropriate sample size is 59, the sample of this study consists of 59 heads of departments, managers and senior management, and 59 questionnaires were distributed to the respondents.

4.2 Questionnaire

Researchers adopted questionnaires from previous studies to test the current study, the questionnaire has been validated to be ideal for measurement by previous researchers, and the questionnaire has been reviewed (Gill et al. 2011; Sadeg hian and Pihie 2012; Murphy & Ensher 2008; Flynn & Staw 2004; Hinkin & Tracey 1999; Judge and Piccolo 2004), the questionnaire consisted of two main parts, the first section contained five questions about each participant's personal information, Participant's age, gender, education level, and job title. Part 2 of the survey consisted of 121 items for all variables, this study examines three leadership styles (transformative leadership, transactional leadership, and democratic leadership) as independent factors and crisis management as a dependent factor, participants were 1 = Strongly Disagree, 5 = Strongly Agree.

4.3 Analysis and Results

As already mentioned, a total of 59 participants participated in the survey, respondents were asked to rate the importance of each item on a scale of 1 to 5, All data were analyzed using SPSS version Table 1. Reliability Test.

Table 1. Reliability test

Factor	Cornbrash's Alpha	No of items
Transformational leadership	0.76	37
Transactional leadership	0.74	26
Democratic leadership	0.78	28
Crisis management	0.71	30

Table 1 Shows a reliability test for three independent variables and one dependent variable, alpha for transformational leadership style = 0.760 > 0.7 (Kothari 2004), this represents 37 items for transformational leadership style, means that was used, reliable, transactional leadership style alpha = 0.74 this means that the 26 items of the transactional leadership style are authoritative, alpha for democratic leadership style = 0.780, meaning 28 items used in democratic leadership Style are trustworthy, alpha for crisis management = 0.710 30 items used in crisis management are trustworthy means.

Table 2 he graph below illustrates the correlations between the three independent and dependent variables. Crisis management was found to be significantly correlated with transformational leadership style ($p = 0.046$ ($p < 0.05$)) in terms of the strength of the linear relationship, and is intermediate between transformational Leadership and crisis management. Regarding the strength of the linear association between transactional leadership and crisis management, the democratic leadership style does not significantly correlate with crisis management, whereas the transactional leadership style does ($p = 0.0452$ ($p < 0.05$)) ($p = .0765$), $p > 0.05$.

Table 2. Pearson product-moment correlations between independent and dependent variables

Variables	1	2	3	4
(1)Crisis Management	1			
(2) Transformational leadership	0.0452	1		
(3)Transactional leadership	0.0464	0.486	1	
(4) Democratic Leadership	0.0765	0.0187	0.010	1

5 Conclusion

The purpose of this study was to examine the role of administrative leadership in crisis management in Jordanian ministry of planning and international cooperation the relationship this study focused on the relationship between transformational leadership, transactional and democratic and crisis management in the Jordanian ministry of planning, the survey sample includes 67 department heads, managers, and senior executives, a total of 59 questionnaires were distributed to the respondents, this research shows that those with a transformative leadership style can take advantage of it, the results of this study are consistent with those of Ali and Anwar (2021).

A crisis is a chaotic situation that an organization or society encounters and can cause chaos. There is no way to completely prevent it. It is impossible to predict the outcome, or at least it is impossible to predict, the measures to prepare for the crisis and the results to be achieved, the necessary precautions to be taken, the early warning

system to be developed, the minimum to overcome the crisis at the cost of The Damage of Managing Disruption Meanwhile, while the situation created by the crisis, other crisis measures taken have been staff layoffs, cuts in spending, and halted investment, certain organizations are ignoring the future, and went too far with such an application.

Meanwhile, given the temporary nature of the crisis, certain other organizations are preparing for the recovery phase without stopping investments. The most important point that needs to be mentioned regarding this topic is that the easiest way to overcome a crisis is to take the necessary precautions in your organization before the crisis occurs.

References

- Abdullah, M.S., Toycan, M., Anwar, K.: The cost readiness of implementing e-learning. *Custos eagronegocio* on line **13**(2), 156–175 (2017)
- Abuhashesh, M.Y., Alshurideh, M.T., Sumadi, M.: The effect of culture on customers' attitudes toward Facebook advertising: the moderating role of gender. *Rev. Int. Bus. Strat.* **31**(3), 416–437 (2021)
- Agha, K., M. Alzoubi, H., Alshurideh, M.T.: Measuring reliability and validity instruments of technologically driven cognitive intrusion towards work-life balance. In: Hassanien, A.E., et al. (eds.) AICV 2021. AISC, vol. 1377, pp. 601–614. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_53
- Ahmed, S.A., et al.: Students' attitudes towards learning English in the Kurdistan region of Iraq. *Int. J. English Literat. Soc. Sci.* **6**(3), 072–087 (2021). <https://doi.org/10.22161/ijels.63.11>
- Akoi, S., et al.: Elaborating the characteristics that affect buyers in online shopping: the case of generation Z Girls in Kurdistan Region of Iraq. *Black Sea J. Manage. Market.* **2**(2), 42–64 (2021). <https://doi.org/10.47299/bsjmm.v2i2.74>
- AlKurdi, B., Alshurideh, M., Alafaishat, T.: Employee retention and organizational performance: evidence from banking industry. *Manage. Sci. Lett.* **10**(16), 3981–3990 (2020)
- Al Shraah, A., Abu-Rumman, A., Alqhaiwi, L., Alshurideh, M.T.: The role of AACSB accreditation in students' leadership motivation and students' citizenship motivation: business education perspective. *J. Appl. Res. High. Educ.* (ahead-of-print) (2022)
- Alameeri, K., Alshurideh, M., Kurdi, B.A., Salloum, S.A.: The effect of work environment happiness on employee leadership. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020*, pp. 668–680. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-58669-0_60
- Manal Ahmad, A.: *The Distinguished Leader and the Secrets of Leadership Creativity*, 1st edn. The Arab Group for Training and Publishing, Cairo (2015)
- Al-bawaia, E., Alshurideh, M., Obeidat, B., Masa'deh, R.: The impact of corporate culture and employee motivation on organization effectiveness in Jordanian banking sector. *Acad. Strat. Manage. J.* **21**(Special Issue 2), 1–18 (2022)
- Al-Dhuhouri, F.S., Alshurideh, M., Kurdi, B.A., Salloum, S.A.: Enhancing our understanding of the relationship between leadership, team characteristics, emotional intelligence and their effect on team performance: a critical review. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) AISI 2020. AISC, vol. 1261, pp. 644–655. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-58669-0_58
- Ali, B.J., Anwar, G.: Marketing strategy: pricing strategies and its influence on consumer purchasing decision. *Int. J. Rural Develop. Environ. Health Res.* **5**(2), 26–39 (2021). <https://doi.org/10.22161/ijreh.5.2>

- AlShehhi, H., Alshurideh, M., Kurdi, B.A., Salloum, S.A.: The impact of ethical leadership on employees performance: a systematic review. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020*, pp. 417–426. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-58669-0_38
- Alshurideh, M.: Does electronic customer relationship management (E-CRM) affect service quality at private hospitals in Jordan? *Uncertain Supply Chain Manage.* **10**(2), 325–332 (2022)
- Alshurideh, M., Al Kurdi, B., Alshurideh, H., Alkurdi, S.: The impact of leadership styles on work from home (WFH) during the COVID-19 situation. *Int. J. Theory of Organ. Pract.* **2**(1), 115–128 (2022)
- Alshurideh, M., Kurdi, B., Alzoubi, H., Obeidat, B., Hamadneh, S., Ahmad, A.: The influence of supply chain partners' integrations on organizational performance: the moderating role of trust. *Uncertain Supply Chain Manage.* **10**(4), 1191–1202 (2022)
- Andover, V., Ali, B.J., Ali, S.A.: Rainwater for water scarcity management: an experience of Woldi a University (Ethiopia). *J. Bus. Econ. Environ. Stud.* **10**(4), 29–34 (2020). <https://doi.org/10.13106/jbees.2020.vol10.no4.29>
- Anwar, G., Shukur, I.: the impact of recruitment and selection on job satisfaction: evidence from private school in Erbil. *Int. J. Soc. Sci. Educ. Stud.* **1**(3), 4–13 (2015)
- Anwar, G., Surarchith, N.K.: Factors affecting shoppers' behavior in Erbil, Kurdistan-Iraq. *Int. J. Soc. Sci. Educ. Stud.* **1**(4), 10 (2015)
- Anwar, K.: The role of effective leadership in crisis management: study of private companies in Kurdistan. *Qalaai Zanist Sci. Journal* **2**(4), 326–338 (2017)
- Anwar, K., Balcioğlu, H.: The relationship between transformational leadership characteristics and effectiveness: a case study of construction companies in Erbil. *Int. J. Sci. Technol. Manage.* **5**(2), 250–256 (2016)
- Anwar, K., Climis, R.: Analyzing the relationship between types of advertisement and customer choice: a study of retailer stores in Erbil. *Int. J. Account. Bus. Soc.* **25**(2), 43–52 (2017)
- Anwar, K., Qadir, G.H.: A study of the relationship between work engagement and job satisfaction in private companies in Kurdistan. *Int. J. Adv. Eng. Manage. Sci.* **3**(12), 239944 (2017)
- Aziz, H.M., et al.: Employee commitment: the relationship between employee commitment and job satisfaction. *J. Human. Educ. Develop.* **3**(3), 54–66 (2021). <https://doi.org/10.22161/jhed.3.3.6>
- Bass, B.M., Avolio, B.J.: Transformational leadership: a response to critiques. In: Chemers, M.M. (eds.) *Leadership Theory and Research: Perspectives and Directions*, pp. 49–80. Academic Press, San DIEGO (2004)
- Bass, B.M.: From transactional to transformational leadership: learning to share the vision *Organizational Dynamics*, (Winter), pp. 19–31 (1990)
- Berson, Y., Avolio, B.J.: Transformational leadership and the dissemination of organizational goals: a case study of telecommunication firm. *Leaders. Q.* **15**, 625–646 (2004)
- Burns, J.M.: *Leadership*. Harper & Row, New York (1978)
- Crawford, C.: *Effects of transformation leadership and organization* (2005)
- Demir, A., Maroof, L., Sabbah Khan, N.U., Ali, B.J.: The role of E-service quality in shaping online meeting platforms: a case study from higher education sector. *J. Appl. Res. High. Educ.* (2020). <https://doi.org/10.1108/JARHE-08-2020-0253>
- Faraj, K.M., Faeq, D.K., Abdulla, D.F., Ali, B.J., Sadq, Z.M.: Total quality management and hotel employee creative performance: the mediation role of job embeddedness. *J. Contemp. Issues Bus. Govern.* **27**(1), 3838–3855 (2021). <https://doi.org/10.47750/cibg.2021.27.01.001>
- Hamza, P.A., et al.: Recruitment and selection: the relationship between recruitment and selection with organizational performance. *Int. J. Eng. Bus. Manage.* **5**(3), 1–13 (2021). <https://doi.org/10.22161/ijebm.5.3.1>

- Alsamaray, H.A.S.: Applied Science University, P.O. Box 5055 Jufair, Kingdom of Bahrain E-mail of the corresponding author: Alsamaray@yahoo.co.uk p38 (2014)
- Ismael, N.B., et al.: The role of training and development on organizational effectiveness. *Int. J. Eng. Bus. Manage.* **5**(3), 15–24 (2021). https://doi.org/10.22161/ije_bm.5.3.3
- Jamal Ali, B., Akoi, S., Fadel Saleh, P., Zhilwan, S.: Factors shaping customer satisfaction with residential flats: evidence from Sulaymaniyah City. *BlackSea J. Manage. Market.* **2**(2), 1–12 (2021). <https://doi.org/10.47299/bsjmm.v2i2.69>
- Mughlezi, A.R.: Designing model of transformational leadership in public organizations of Iran, *Quart. J. Manag. Knowl.* **621**–18 (2003) (InPersian)
- Spreitzer, G.M., Perttial, K.H., Xin, K.: Traditionally matters: an examination of the effectiveness of transformational leadership in the United States and Taiwan. *J. Organize. Behave.* **26**, 208–209 (2005)
- Odeh, R.B.M., Obeidat, B.Y., Jaradat, M.O., Alshurideh, M.T.: The transformational leadership role in achieving organizational resilience through adaptive cultures: the case of Dubai service sector. *Int. J. Product. Perform. Manage.* (2021)
- Sabir, B.Y., et al.: Administrative decentralization: the transfer of competency from the ministry of education to general directorates. *Int. J. Rural Develop. Environ. Health Res.* **5**(3), 1–13 (2021). <https://doi.org/10.22161/ijreh.5.3.1>
- Sorguli, S., et al.: Innovation: Knowledge Management in the Innovating Industries , vol. 6, issue 3, pp. 10–23 (2021). <https://doi.org/10.22161/eec.63.2>
- Wirawan, Kepemimpinan: Teori, Psikologi, Perilaku Oganisasi, Aplikasi dan Penelitian (Jakarta: PT Raja GrafindoPersada, 2014), p. 6 (2014). www.Emeraldinsight.com Leadership theory and research: Perspectives and directions



Remote Working in the COVID-19 Era

Randa Abu Hamour¹, Areeg Alfouri²(✉), and Muhammad Alshurideh^{3,4} 

¹ Al Balqa Applied University, As-Salt, Jordan
Randa.nayfe@yahoo.com

² Faculty of Medicine, Al Balqa Applied University, As-Salt, Jordan
areeg.alfouri@bau.edu.jo

³ Department of Marketing, School of Business, The University of Jordan, Amman 11942, Jordan

m.alshurideh@ju.edu.jo, malshurideh@sharjah.ac.ae

⁴ Department of Management, College of Business, University of Sharjah, Sharjah 27272, United Arab Emirates

Abstract. The COVID-19 pandemic is the largest virtual experience for online work and employee engagement in the world. Millions of employees have been required to spend extended periods of time working on digital platforms such as Zoom, Microsoft Team, and Skype. However, the nature of this work has resulted in a number of negative implications, such as lengthy work hours, virtual meetings, and the constant pressure to check and respond to business-related emails. In addition, several employees felt obligated to stay online to demonstrate that they were not avoiding their work responsibilities, which contributed to their psychological stress. In this research, a questionnaire was utilized to assess remote work throughout the COVID-19 pandemic's spread. The questionnaire consisted of four axes: work intensification, employee engagement and online presence, adapting to new working techniques, and demographic aspects of the respondents. According to the results, the abrupt change from physical to online work patterns increased work intensity, online presence, and job instability. We also observe that the transition to new methods of working from home is proceeding well.

Keywords: Remote work · The COVID-19 pandemic · Virtual meetings

1 Introduction

In December 2019, there was an outbreak of acute pneumonia of unknown causes, and many cases were recorded in Wuhan, Hubei Province, China. Later, the causative agent of this infection was identified as SARS-CoV-2 (COVID-19), which was classified by the World Health Organization on March 11, 2020, as a pandemic (WHO 2022).

Coronavirus is a membrane-bound, single-stranded RNA virus with a diameter of approximately 0.1 μm (Majumder and Majumder 2021). The virus is transmitted via droplets, such as those produced by coughing, direct contact with an infected individual, or touching contaminated hands or surfaces. The COVID-19 pandemic is the greatest threat to humanity since World War II and a current public health disaster. The virus first

arrived in Asia near the conclusion of 2019 before spreading to the rest of the world. The pandemic is more than a health crisis affecting numerous countries; it has the potential to have devastating social, economic, and political repercussions that might have long-lasting effects. Every day, hundreds of residents lose their employment with no way of knowing when life will return to normal (Abed 2022).

The effects of the COVID-19 pandemic shocked society, as it are inevitable. The entire behavior of the economy, society, work, and lifestyles has changed (Alshurideh et al. 2021a; Amarneh et al. 2021; Leo et al. 2021). Consumer's behavior has also changed by destroying existing habits or creating new ones (Ahmad et al. 2021; Alameeri et al. 2021; Al-Dmour et al. 2021a). Therefore, leading institutions had to shift towards a digital society, innovate, and take action for survival and sustainability (Harahsheh et al. 2021; Khasawneh et al. 2021a and b).

Then, most countries started a gradual lockdown to control the outbreak by isolating the cases and tracing the contacts of the patients (Adisa et al. 2021). The Corona pandemic changed the work environment, leaving many offices empty and most employees directed to work from home as a way to limit the spread of the virus (Adnan & Anwar 2020; Al-Dmour et al. 2021b). While many thought that it would be temporary, this closure lasted for more than a year and compelled many organizations with new work options, especially flexible and remote work (Akour et al. 2021; Alshurideh et al. 2021b; Mitchell 2021).

Hence, this experience created uncertainty in the expectations of employees and employers regarding the extent to which remote work can be effectively implemented, and managed (Matli 2020; Taryam et al. 2020). Although remote work is not new, the pandemic has revealed its pros, cons, and shortcomings (Aljumah et al. 2021; Nuseir et al. 2021).

Coronavirus has become a worldwide problem requiring the commitment of authorities, legislators, academics, and scientists in order to handle its numerous dimensions. Recent research reveals that the COVID-19 virus creates panic, uncertainty, and anxiety around the planet. The issue of health became a cause of sufferings that disrupted family and work life. Furthermore, the pandemic has led to economic destabilization. Concerning the severity and duration of this infection, there is still considerable uncertainty. The World Health Organization (WHO) pronounced the disease to be a pandemic, and in effort to contain the epidemic and decrease its spread, a number of nations enacted restrictions on human meetings. In turn, this affected schools, universities, and colleges (Alshurideh et al. 2019a; Al Kurdi et al. 2020a; Hilburg et al. 2020; Shah et al. 2021). Most profit or non-profit organizations had to reduce their departments or close them entirely.

One can imagine that about 1.6 billion children in 195 countries were unable to use their classrooms during the peak of the COVID-19 virus (Daroedono et al. 2020). If the closure of these educational institutions is extended for a longer period, there is a high possibility of losing educational opportunities, a potential loss of human capital, and a decrease in the economic opportunities in the long term. Hence comes the idea of searching for alternative methods in the educational process to access education in addition to self-education to reduce the impact of closures and bans by using technology tools and ICT communication technology, such as Team, Zoom and several similar

softwares (Alshurideh et al. 2019b; Al Kurdi et al. 2020b; Sindiani et al. 2020; Sultan et al. 2021; Alshurideh et al. 2023). This beside the online courses and videos uploaded to the educational platforms.

Despite the benefits of remote work, the various organizational sectors did not fully do so, questioning the strengths, acceptability, and the extent of achievement and effectiveness of remote work (Alhomdy et al. 2021). Hence, some researchers argued that remote work discourages employees and encourages laziness, while others emphasized that the pandemic led to the emergence of a flexible work style. Still, others emphasized that it is not valid in all sectors and organizations, whether economic, service, or in the construction and manufacturing sector.

Most studies during the pandemic focused on flexibility in the workplace (Gavrila et al. 2021) and the consequences and the “dark side” of remote work, especially the psychological state of employees and workers in industrial fields and the reactions of managers and their satisfaction with this process.

The Covid-19 pandemic is different from anything we have seen in recent times. It has had devastating effects on people’s mental health and well-being and has caused serious problems for families and communities all over the world. Large numbers of people have lost their jobs and source of income. Others, especially in developed economies and countries, have been forced out of work and became dependent on government support. It has also taken a heavy toll on the global economy since the world has seen a steady decline in revenue generation as organizations continue to cut costs (Gavrila et al. 2021), reduce employee benefits, and invest less in human capital spending. The world has faced several social and economic challenges with national lockdowns, stay-at-home orders and other social distancing measures to curb the spread of the virus. While these restrictions have reduced the transmission of the disease, it has also forced many employees to make a sudden shift to remote work.

These changes in the work atmosphere had several far-reaching consequences for both individuals and organizations, as several questions emerged, including: is remote work really beneficial for employees? Does it offer flexibility to people working remotely? Are employees able to achieve cost savings on travel, spend less time transporting, and strike balance between their work and family-related duties?

Despite this, there are many limitations to remote work, including communication between employees, lack of motivation to work, loss of teamwork, problems related to data security, and difficulty in monitoring performance (Alshurideh et al. 2014; Alyammahi et al. 2020; Alrawabdeh et al. 2022). These issues present real problems for many employees and ultimately employers, not to mention the conflicting demands between work and family as a major cause for stress, drains social life and individual resources, and negatively affects work engagement.

Working from home is a kind of conflict between the tasks assigned to the employee and his family, especially the tasks assigned to female employees, such as housework and childcare responsibilities besides technical problems, such as poor access to digital technologies, small workspaces at home and distracting noises inside the employee’s house or from neighbors. This is in addition to psychological stress that can deplete limited social and personal resources and negatively affect work engagement.

2 Remote Working and Companies' Solutions

The term remote working (RW) first spread in the 1970s, expressed as “work from home”, which is individuals practice of working from their homes instead of a central workplace, after telecommunications networks had become available to employees in their homes.

Then various technologies and information and communication technology (ICT) were used to communicate in the work environment as an alternative to travel beside the use of remote work technologies and virtual technologies (Uwizeyimana 2022). Thus, the COVID-19 pandemic has forced employers to adopt a new approach, which is flexibility in the workplace and thus benefit from the positive effects of remote work.

In addition, the shift to remote work arose in the wake of the pandemic and thus the challenges facing organizations emerged, including quickly establishing the appropriate infrastructure needed for employees to work from home (digital assistive technologies, software, physical equipment, and organizational processes to facilitate effective Distance working) (Vahdat 2022).

From another point of view, employees see that remote work increases work flexibility, increases job independence, increases productivity, reduces business costs, reduces employee turnover, enhances work-life balance, improves job satisfaction, reduces mobility costs, and increases participation in work. While employers see some consequences of remote work, including limited communication, employee isolation, decreased dedication and commitment, and employers' loss of control over employees' work procedures. This is in addition to the unsupportive work culture, health and safety policies, and management concerns about trust and performance uncertainty that also limit the effectiveness of remote work.

Although employees have the legal right to request flexible working arrangements, organizations are not authorized to agree to such requests. A survey of the Confederation of Trade Unions in 2019 found that one of every three applications for flexible working was rejected.

The literature also identified the most important determinants of employee participation, as this process is affected by many job characteristics that enable them to practice their work, including independence, mission identity, diversity of skills, importance of the task, and inspiring employees to make them more effective and enhancing their skills in order to improve the overall quality of what they do. In other words, the participation of employees centers around the main aspects of the jobs or tasks assigned to them, the degree of the tasks' importance, and the senior management and employers attention to them.

The participation of employees and remote work during the pandemic was socially linked to isolation, long working hours, and conflicting work and family demands. This isolation was associated with tension, psychological pressure, and mental stress among some employees. The basic hypothesis for employees is that individuals strive to keep resources (money, shelter, and basic needs), social needs (personal support and mutual respect), personal needs (positive self-esteem and optimism), and factors to improve well-being and enhance the individual's feeling.

Hence, it can be said that a large number of employees have increased their work intensity during the Covid-19 pandemic, as the pandemic forced them to stay at home

and work remotely. They were also required to take more tasks and duties towards their families (such as studying from home and childcare responsibilities) and deal with the sudden shift to remote work. All these changes led to an increase in the effort exerted by the employees. Many employees have seen the dramatic increase in phone calls, emails, online meetings, training, and long and stressful working hours. Therefore, a large percentage of employees emerged who were exposed to job loss and layoffs.

It can be also argued that the outbreak of Covid-19 is the largest virtual experience in the world for online work and employee participation, as it forced millions of employees to work for long periods of time on digital platforms, such as Zoom, Microsoft Team and Skype. However, the nature of this work had several negative consequences, including never ending long work hours, virtual meetings, and the constant pressure to check/reply work emails or messages outside working hours. Moreover, some employees felt compelled to stay online just to prove that they were not shirking work responsibilities, which added to the psychological pressure on them. Some employees felt that they were just robots in front of computers and that they are prisoners to these technologies. The virtual workplace has replaced the traditional office; this kind of virtual work is not normal (just virtual meetings, virtual teaching, virtual collaboration, virtual training courses, online student support, online supervision), as described by some employees. In this experience, you feel that you are always being watched in this virtual world and you have to be logged in at all times or your manager will consider you slacking.

Multiple web meeting services, such as Google Meet, Zoom, Microsoft Teams, WebEx, etc., reported a revenue gain of more than 100 percent as a result of a sudden increase in the use of their virtual digital technologies. Moreover, the lack of job security and employee engagement has contributed to poor job engagement. Also, due to the massive social and economic cost of the pandemic, many employers have taken some cost-cutting measures aimed at reducing employee benefits, reducing business operations, and restructuring employee jobs, creating an atmosphere of uncertainty in the working conditions of employees. One of the employees described that period as terrible and very tense, and no one except the main employees (doctors and nurses) felt job security, while many employees felt fear of unemployment or layoffs.

The functioning, adaptability, and sustainability of the supply chain were impacted by the pandemic's spread. In addition, employees have been compelled to work from home, which has caused a disruption in the extent of emotional distress, resulting in a reduction in productivity and, ultimately, a decline in income. Furthermore, limitations such as physical and social separation, lockdowns, and travel bans provided a fundamental barrier for enterprises in directing services and contacting clients.

The majority of service businesses, particularly those in the hospitality and tourism sectors are having difficulty keeping their services operational, while others are going to shut their doors, posing real risks to the well-being of their employees and generating a number of societal and financial challenges. The spread of the Corona virus precipitated an unparalleled health and economic crisis. Notable here is that future leaders must focus on comprehending the magnitude of such crises and thoroughly examining their societal, financial, and technical effects. In 2020, worldwide supply and demand failed, and the volume of merchandise commerce in the global economy declined by 17.7 percent in May 2020 when compared to the same month in 2019. This is in addition to a decline

in domestic demand and investments in commercial activities, a decline in tourism and business travel, a decrease in production and trade activities, disruptions in demand and supply, effects on health such as increased morbidity and mortality, and changes in healthcare expenditures. Inevitably, service organizations; including banks, and the hospitality and tourism industry have suffered hard consequences (Gössling et al. 2021). Hence, more jobs are at risk if travel restrictions and quarantine practices continue.

A global shock known as COVID-19 had an impact on everyone's lives and livelihoods and resulted in new challenges, particularly for commercial enterprises and supply and demand, interruptions in the supply of products and services, and decreased worker availability, productivity, and efficiency. It also had an impact on consumer and household demand. Job cuts and wage cutbacks as a result of illness, quarantine, and unemployment have negatively impacted consumption and organizational investments. Various measures have been put into place by nations around the world, including personal counseling, cleanliness instructions, progressive social and physical segregation, work from home policies, travel limitations, and lockdown. The majority of the organizations lost their skilled and experienced workers as a result of the pandemic, which had a significant negative impact on social capital. This reduced the labor pool, contributed to poverty due to the loss of the breadwinner's life and lower productivity, and eventually reduced employee wages.

COVID-19 has compelled businesses to digitize, permit more work to be performed from home, and adopt a variety of technological and digital practices; spending on digital transformation has surged, and organizations around the world are paying more money to purchase the most recent technology. Technology will play a significant role in the present and future, and the use of digital technology can enable the workforce to work from home and complete their professional responsibilities using smartphones, tablets, laptops, and desktop computers to keep the business afloat during the current economic crisis.

Moreover, companies that have designed their solutions to use the full potential of the latest technologies, such as cloud computing, access to digital payment methods, virtual meetings, and video conferencing, can create resilience in the face of future threats by implementing future business concepts and practices that are already under exploration. The sudden increase in online activities has also increased economic costs and cybersecurity risks due to more people working remotely. In addition to training employees with less knowledge, high reliance on technology, and data integration problems. There are also other challenges like the decline in sales/revenue, decline in manufacturing, customer issues, decline in global trade, unfamiliarity with online applications, inadequate infrastructure, and difficulty communicating with customers or suppliers in a digitally connected environment. However, companies have faced several obstacles and difficulties in responding to this pandemic, the challenge of working remotely and keeping pace with the new phenomenon. For example, several companies in the field of cloud computing invested in providing infrastructure shortly before the outbreak of the pandemic, which helped them to overcome that difficult stage. Cloud computing helped employees and co-workers collaborate and communicate securely and share resources and knowledge. In addition, during the COVID-19 period, cloud technology played a major role in fighting the pandemic in the sense that the salvation of governments and

organizations in many areas of life, education, health, industry, communications and monitoring was through remote work and more information exchange.

Companies and factories have become more dependent on technology, information, and communication companies. One of the chief executive bloggers of Google Cloud spoke about the increase in demand for Google Cloud services, and the daily growth rate of the Meet application has exceeded 60%. As a result, the daily use is 25 times higher than it was in January of the previous year. In a positive development, Google revealed that advanced features in Google Meet are now free for both G Suite and G Suite for Education customers around the world. Most technology and communications companies have made it clear that most digital devices, i.e. smart phones, tablets, and laptops, enable any employee to work remotely, but their effectiveness, ability to work, and speed of their internet and cloud connection vary through cloud computing applications.

Here it is worth noting that most companies and institutions resort to digitization in their work. Digitization refers to the purchase of new technological assets or the upgrade of their existing assets; these assets can be devices, such as computers, servers, network devices, and sensors, or software such as operating systems, applications, or even digital content or services. Here, a new concept appears, which is cloud computing and its extensive use. Some industries have shifted their business towards benefiting from cloud computing services as follows:

1. **Software as a Service (SAAS):** It is a software deployment model in which a person or organization specifies the services they want at the time of demand (on-demand).
2. **Platform as a Service (PAAS):** It means the delivery or the computing system to developers, as it facilitates the deployment and spread of applications without any cost or difficulty, whether in the purchase of basic hardware or software.
3. **Infrastructure as a Service (IAAS):** providing infrastructure for technology, such as network, storage, and account, represented by a virtual service. Customers may also control deployed operating systems, storage, and applications and possibly have limited control over specific networking components, e.g., host firewalls.

Other indicators can also be added, such as artificial intelligence, machine learning, big data analytics, 5G Internet, Internet of Things, object sensors, super automation, robotics, cybersecurity, 3D printing, and even some technical devices, such as drones.

It is worth noting that Italy was one of the countries most affected by the Corona pandemic, and it was one of the least responsive countries to remote work before the pandemic, which imposed several challenges and difficulties in coping with the pandemic.

The most important positive results of remote work include increased productivity, providing greater independence, and improving flexibility (Barabaschi et al. 2022). Remote work is meant for the ability to work from various locations without restrictions due to working hours and risks related to displacement and crowding, in order to improve efficiency and performance.

This pandemic led to the generation of a new type of income for individuals to be able to pay payments or recover some investments, leading to some governments around the world pumping money into economic activities through means of active policies to

mitigate the effects of the epidemic, including some subsidies targeting the workforce, with their national health care systems providing full insurance in the case of exposure to COVID-19. Other benefits targeted families at risk of poverty by providing them with a basic income for a short period. From a corporate perspective, governments have targeted tax exemptions and temporary employee subsidies for organizations whose business was disrupted by the pandemic and forced to completely close or significantly reduce their commercial activity, thus avoiding mass layoffs, especially in specific sectors, such as the service sector, hospitality, tourism, and leisure services.

During the peak of the COVID-19 pandemic, businesses of all types, but especially those in the food industry, were compelled to take entrepreneurial and inventive measures in order to respond to the changing marketplace and seize new opportunities. Organizations might use this time to rethink their business models, make a strategic pivot, or fully embrace digital transformation in order to better meet the evolving demands of their customers. Here, we see how the flexibility of companies has helped them grow and succeed. There was a spike in internet use during the lockdown, suggesting that businesses were looking for ways to keep operating and were ultimately deciding to digitize in order to maintain a presence online. Therefore, digital transformation is now more important than ever for businesses to continue operating, and organizations must work to develop a mechanism based on analyzing specific indicators that can objectively measure growth or decline in rates of entrepreneurship, innovation, or digitization among companies experiencing extraordinary events like the recent COVID-19 pandemic. In terms of long-term planning, capital investment, financing, economic policies, and available lines of credit, this can allow for quicker, more informed, and more drastic action.

With the spread of cloud computing, companies or organizations that adopted this system have a certain level of concern about the security of their data and files since they do not know where and how their data is secured with the cloud provider. That is why public and private organizations have invested millions in security products and incident response personnel but have not tested the efficiency of these guarantees and the people who use actual attack techniques. In addition, they generally lack a comprehensive approach to understanding and operationalizing cyber threat intelligence; therefore, their defensive measures are weak and tainted by a large degree of opacity and high security levels. However, cloud providers make sure that they monitor customer data and files at all times using the latest security policies and services through the shared responsibility model followed on the cloud. Thus, it is necessary for the organization to monitor, identify, and address any potential threats and misconfigurations on the cloud resources provided. However, data breaches, identity management, and incorrect access methods are important factors to consider before moving to cloud adoption.

When organizations adopt cloud computing, they get eternal benefits of cloud adoption including high availability, expandability, business continuity, fault tolerance, disaster recovery, automatic software updates, resilience, etc. With the passage of time, a lot of new features and services have emerged with the growing interest in adoption of cloud computing by various organizations as Amazon Web services originally emerged as a sustainable cloud. Moreover, several technology companies have emerged that provide cloud services with different features and advantages, such as Microsoft's Azure, Google Cloud, IBM Cloud, and Oracle.

It is critical to decide how the data will be stored and what security features will be applied to it before giving access to the cloud provider. How secure it is depends on the level of confidential data that you currently possess. The procedures for ensuring the security of data and how to access it must be clearly defined and communicated to the organization by the service provider, and the data transmitted or stored must be encrypted and the level of encryption can be mutually determined.

Data security in the cloud world is more complex than data security in local or traditional information systems because the data is scattered in different places on different devices. It is worth noting that organizations must know the location of the data due to the existence of privacy and security laws that apply to data in many countries. Sometimes data is required to be located on more than one continent to make it highly available but local privacy and security laws cannot be ignored.

3 Evaluation Method: A Questionnaire to Evaluate Remote Work During the COVID-19 Pandemic

In this study, a questionnaire was followed to evaluate remote work during the COVID-19 pandemic, and the questionnaire was divided into four axes, namely:

- Work intensification (Category 1).
- Employee Engagement and Online presence (Category 2).
- Adaptation to new working approaches (Category 3).
- Demographic factors of the respondents (Category No. 4).

The questionnaire was distributed to all employees of the Faculty of Medicine, Al-Balqa'a Applied University, (the administrative body), and the total number of participants was 50 participants. Table 1 shows the question and matching options.

Table 1. Question and matching options

Option	Value
Strongly disagree	0.00
Disagree	0.25
Average	0.50
Agree	0.75
Strongly agree	1.00

A value is assigned to respond to each question, and then accrual points are collected according to the four categories described above, according to the following formula:

$$Q = \frac{\text{total of remote work index for a category in each question}}{\text{Number of questions}}$$

The following table shows the points for measuring remote work, which are as follows:

Points (x)	Level of remote work
$0 < = x < = 0.2$	Bad
$0.2 < x < = 0.4$	Weak
$0.4 < x < = 0.6$	Average
$0.6 < x < = 0.8$	Good
$0.8 < x < = 1.0$	Excellent

4 Discussion of the Results

4.1 Work Intensification, Employee Engagement, and Online Presence

All participants reported that their workloads have intensified since the Covid-19 lockdown, and stay-at-home orders forced them to work from home and were required to take on more tasks despite having to manage additional family-related duties (such as homeschooling and childcare responsibilities) and deal with sudden shift to work remotely. These changes increased the amount of effort participants had to invest in their jobs, which ultimately lowered levels of work engagement. Participants also reported an increase in stress, especially in March and June 2020 (the period of national lockdown in Jordan).

Category	Points	Remote work index
Work intensification	0.67	Good

4.2 Adapting to New Working Approaches

The Covid-19 outbreak is often described as the world’s largest virtual experience of remote working; millions of employees around the world have been forced to work for long periods of time on digital platforms, such as Zoom, Microsoft Team, and Skype, and work from home and not go to their workplace. Employees were forced to attend virtual meetings and use multiple systems, such as Face Time, Skype, Zoom, to always be available online and respond to work-related tasks and the constant pressure to check/respond to work-related emails or messages outside of working hours. Moreover, some participants felt compelled to remain visible online just to prove that they were not shirking work responsibilities while working from home. This greatly caused fatigue and monotony as a result of the great commitment in front of the computer screen. Table 2 describes the demographic factors of the respondents.

Category	Points	Remote work index
Adaptation to new work approaches	0.70	Good

Table 2. Demographic factors of the respondents

		Percentage
Gender	Male	20%
	Female	80%
Job nature	Administration staff	
	Manager	80%
Age	21–30 years	20%
	31–40 years	
	41–50 years	
	Above 50	
Education	Under high school	2
	High school	2
	Intermediate diploma	21
	Bachelors	20
	Masters	3
	PHD	2

5 Conclusion

The purpose of our study was to explore remote working and remote employee engagement in the context of Covid-19. Based on the results of the questionnaire, our results revealed that the sudden transition from physical to online work increased work intensity, online presence, and job insecurity. We also noticed that the adjustment to new ways of working from home was going well. Further, the exposure to stress factors among employees led to psychological stress and poor levels of participation at work.

References

- Abed, S.S.: A literature review exploring the role of technology in business survival during the Covid-19 lockdowns. *Int. J. Organ. Anal.* **30**(5), 1045–1062 (2022)
- Achebak, H., Petetin, H., Quijal-Zamorano, M., Bowdalo, D., Pérez García-Pando, C., Ballester, J.: Trade-offs between short-term mortality attributable to NO₂ and O₃ changes during the COVID-19 lockdown across major Spanish cities. *Environ. Pollut.* **1**(286), 117220 (2021)

- Adekoya, O.D., Adisa, T.A., Aiyenitaju, O.: Going forward: remote working in the post-COVID-19 era. *Empl. Relat.* **44**(6), 1410–1427 (2022)
- Adisa, T.A., Ogbonnaya, C., Adekoya, O.D.: Remote working and employee engagement: a qualitative study of British workers during the pandemic. *Inf. Technol. People ahead-of-print No. ahead-of-print* (2021)
- Adnan, M., Anwar, K.: Online learning amid the COVID-19 pandemic: Students' perspectives. *J. Pedagog. Sociol. Psychol.* **2**(1), 45–51 (2020)
- Ajmal, M.M., Khan, M., Shad, M.K., AlKatheeri, H., Jabeen, F.: Empirical examination of societal, financial and technology-related challenges amid COVID-19 in service supply chains: evidence from emerging market. *Int. J. Logist. Manag. ahead-of-print No. ahead-of-print* (2021)
- Ahmad, A., Alshurideh, M.T., Al Kurdi, B.H., Salloum, S.A.: Factors impacts organization digital transformation and organization decision making during Covid19 pandemic. In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence. SSDC*, vol. 334, pp. 95–106. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67151-8_6
- Akour, I., Alshurideh, M., Al Kurdi, B., Al Ali, A., Salloum, S.: Using machine learning algorithms to predict people's intention to use mobile learning platforms during the COVID-19 pandemic: machine learning approach. *JMIR Med. Educ.* **7**(1), 1–17 (2021)
- Alameeri, K.A., Alshurideh, M.T., Al Kurdi, B.: The effect of covid-19 pandemic on business systems' innovation and entrepreneurship and how to cope with it: a theatrical view. In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence. SSDC*, vol. 334, pp. 275–288. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67151-8_16
- Al-Dmour, A., Al-Dmour, H., Al-Barghuthi, R., Al-Dmour, R., Alshurideh, M.T.: Factors influencing the adoption of e-payment during pandemic outbreak (COVID-19): empirical evidence. In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence. SSDC*, vol. 334, pp. 133–154. Springer, Cham (2021a). https://doi.org/10.1007/978-3-030-67151-8_8
- Al-Dmour, R., AlShaar, F., Al-Dmour, H., Masa'deh, R., Alshurideh, M.T.: The effect of service recovery justices strategies on online customer engagement via the role of "customer satisfaction" during the Covid-19 pandemic: an empirical study. In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence. SSDC*, vol. 334, pp. 325–346. Springer, Cham (2021b). https://doi.org/10.1007/978-3-030-67151-8_19
- Aljumah, A., Nuseir, M.T., Alshurideh, M.T.: The impact of social media marketing communications on consumer response during the COVID-19: does the brand equity of a university matter? In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence. SSDC*, vol. 334, pp. 367–384. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67151-8_21
- Al Kurdi, B., Alshurideh, M., Salloum, S.A.: Investigating a theoretical framework for e-learning technology acceptance. *International Journal of Electrical and Computer Engineering (IJECE)* **10**(6), 6484–6496 (2020a)
- Al Kurdi, B., Alshurideh, M., Salloum, S., Obeidat, Z., Al-dweeri, R.: An empirical investigation into examination of factors influencing university students' behavior towards elearning acceptance using SEM approach. *Int. J. Interact. Mobile Technol.* **14**(2), 19–24 (2020b)
- Al-Balas, M., Al-Balas, H.I., Jaber, H.M., et al.: Distance learning in clinical medical education amid COVID-19 pandemic in Jordan: current situation, challenges, and perspectives. *BMC Med. Educ.* **20**, 341 (2020)
- Al-Fouri, A., Al-Fouri, S.: Cloud computing application in libraries. *ICEPUB-The University of Jordan Library 3. Jordan University- Library.* 193–200p. (2017)

- Alhomdy, S., Thabit, F., Abdulrazzak, F., Haldorai, A., Jagtap, S.: The role of cloud computing technology: A savior to fight the lockdown in COVID 19 crisis, the benefits, characteristics and applications. *Int. J. Intell. Networks* **2**, 166–174 (2021)
- Alrawabdeh, W., Alshurideh, M., Akour, I., Al Kurdi, B.: Impact of change management and communication on business process management. *Int. J. Bus. Anal. Secur.* **1**(1), 117–134 (2022)
- Alshurideh, M.T., Shaltoni, A., Hijawi, D.: Marketing communications role in shaping consumer awareness of cause-related marketing campaigns. *Int. J. Market. Stud.* **6**(2), 163–168 (2014)
- Alshurideh, M., Kurdi, B.A., Salloum, S.A.: Examining the main mobile learning system drivers' effects: A mix empirical examination of both the Expectation-Confirmation Model (ECM) and the Technology Acceptance Model (TAM). In: Hassanien, A.E., Shaalan, K., Tolba, M.F. (eds.) *AISI 2019. AISIC*, vol. 1058, pp. 406–417. Springer, Cham (2020a). https://doi.org/10.1007/978-3-030-31129-2_37
- Alshurideh, M., Salloum, S.A., Al Kurdi, B., Monem, A.A., Shaalan, K.: Understanding the quality determinants that influence the intention to use the mobile learning platforms: a practical study. *Int. J. Interact. Mobile Technol.* **13**(11), 183–157 (2019a)
- Alshurideh, M.T., Hassanien, A.E., Ra'ed Masa'deh: The Effect of Coronavirus Disease (COVID-19) on Business Intelligence. Springer (2021a)
- Alshurideh, M.T., et al: Factors affecting the use of smart mobile examination platforms by universities' postgraduate students during the COVID 19 pandemic: an empirical study. *Informatics*, **8**(2), 1–21 (2021b). Multidisciplinary Digital Publishing Institute
- Alshurideh, M., Abuanezh, A., Kurdi, B., Akour, I., AlHamad, A.: The effect of teaching methods on university students' intention to use online learning: Technology Acceptance Model (TAM) validation and testing. *Int. J. Data Network Sci.* **7**(1), 235–250 (2023)
- Alyammahi, A., Alshurideh, M., Kurdi, B.A., Salloum, S.A.: The impacts of communication ethics on workplace decision making and productivity. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *AISI 2020. AISIC*, vol. 1261, pp. 488–500. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-58669-0_44
- Amarneh, B.M., Alshurideh, M.T., Al, B.H., Kurdi, Z.O.: The impact of COVID-19 on e-learning: advantages and challenges. In: Hassanien, A.E., et al. (eds.) *AICV 2021. AISIC*, vol. 1377, pp. 75–89. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_8
- Arunprasad, P., Dey, C., Jebli, F., Manimuthu, A., El Hathat, Z.: Exploring the remote work challenges in the era of COVID-19 pandemic: review and application model. *Benchmark. Int. J.* **29**(10), 3333–3355 (2022)
- Bakarich, K.M., Marcy, A.S., O'Brien, P.E.: Has the fever left a burn? A study of the impact of COVID-19 remote working arrangements on public accountants' burnout. *Account. Res. J.* **35**(6), 792–814 (2022)
- Barabaschi, B., Barbieri, L., Cantoni, F., Platoni, S., Virtuani, R.: Remote working in Italian SMEs during COVID-19. Learning challenges of a new work organization. *J. Work. Learn.* **34**(6), 497–512 (2022)
- Basilaia, G., Kvavadze, D.: Transition to Online Education in Schools during a SARS-CoV-2 Coronavirus (COVID-19) Pandemic in Georgia. *Pedagog. Res.* **5**, Article No. em0060 (2020)
- Ekarini, D., et al.: The impact of COVID-19 on medical education: our students perception on the practice of long distance learning. *Int. J. Commun. Med. Public Health* (2020). (2020). <https://doi.org/10.18203/2394-6040.ijcmph20202545>
- Gavrila, S., de Lucas Ancillo, A.: COVID-19 as an entrepreneurship, innovation, digitization and digitalization accelerator: spanish Internet domains registration analysis. *British Food J.* **123**(10), 3358–3390 (2021)
- Gössling, S., Scott, D., Michael Hall, C.: Pandemics, tourism and global change: a rapid assessment of COVID-19. *J. Sustain. Tour.* **29**(1), 1–20 (2021)

- Harahsheh, A., Houssien, A., Alshurideh, M., AlMontaser, M.: The Effect of Transformational Leadership on Achieving Effective Decisions in the Presence of Psychological Capital as an Intermediate Variable in Private Jordanian Universities in Light of the Corona Pandemic. *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence*, vol. 334, pp. 221–243 (2021)
- Harel, R.: The impact of COVID-19 on small businesses' performance and innovation. *Global Bus. Rev.* (2021)
- Hellewell, J., et al.: Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Funk S, Eggo RM. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health* **8**(4), e488–e496 (2020). [https://doi.org/10.1016/S2214-109X\(20\)30074-7](https://doi.org/10.1016/S2214-109X(20)30074-7). (Epub 2020 Feb 28. Erratum in: *Lancet Glob Health*. 2020 Mar 5)
- Hilburg, R., Patel, N., Ambruso, S., Biewald, M.A., Farouk, S.S.: Medical education during the coronavirus disease-2019 pandemic: learning from a distance. *Adv. Chronic Kidney Dis.* **27**(5), 412–417 (2020). <https://doi.org/10.1053/j.ackd.2020.05.017>. Epub 2020 Jun 23
- Khalaf, K., El-Kishawi, M., Moufti, M.A., Al, K.S.: Introducing a comprehensive high-stake online exam to final-year dental students during the COVID-19 pandemic and evaluation of its effectiveness. *Med. Educ. Online.* **25**(1), 1826861 (2020)
- Al Khasawneh, M., Abuhashesh, M., Ahmad, A., Alshurideh, M. T., Masa'deh, R.: Determinants of e-word of mouth on social media during COVID-19 outbreaks: an empirical study. In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence*. SSDC, vol. 334, pp. 347–366. Springer, Cham (2021a). https://doi.org/10.1007/978-3-030-67151-8_20
- Al Khasawneh, M., Abuhashesh, M., Ahmad, A., Masa'deh, R., Alshurideh, M. T.: Customers online engagement with social media influencers' content related to COVID 19. In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence*. SSDC, vol. 334, pp. 385–404. Springer, Cham (2021b). https://doi.org/10.1007/978-3-030-67151-8_22
- Kövcses, V.G., Lampert, B., Pongrácz, A., Lörincz, I.: university lecturers' distance learning experiences gained during the COVID-19 pandemic period. In: 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 000325–000330 (2020)
- Kushagra, K., Dhingra, D.S.: Cloud doctrine: impact on cloud adoption in the government organizations of India. *J. Sci. Technol. Policy Manage.* **13**(4), 925–951 (2022)
- Leo, S., Alsharari, N.M., Abbas, J., Alshurideh, M.T.: From offline to online learning: a qualitative study of challenges and opportunities as a response to the COVID-19 pandemic in the UAE higher education context. In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, Rd. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence*. SSDC, vol. 334, pp. 203–217. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67151-8_12
- Majumder, J., Minko, T.: Recent developments on therapeutic and diagnostic approaches for COVID-19. *AAPS J.* **23**(1), 1–22 (2021). <https://doi.org/10.1208/s12248-020-00532-2>
- Matli, W.: The changing work landscape as a result of the Covid-19 pandemic: insights from remote workers life situations in South Africa. *Int. J. Sociol. Soc. Policy* **40**(9/10), 1237–1256 (2020)
- Mitchell, A.: Collaboration technology affordances from virtual collaboration in the time of COVID-19 and post-pandemic strategies. *Inform. Technol. People ahead-of-print No. ahead-of-print* (2021)
- Moszkowicz, D., Duboc, H., Dubertret, C., Roux, D., Bretagnol, F.: Daily medical education for confined students during coronavirus disease 2019 pandemic: a simple videoconference solution. *Clin. Anat.* **33**(6), 927–928 (2020). <https://doi.org/10.1002/ca.23601>. Epub 2020 Apr 22

- Nanath, K., Balasubramanian, S., Shukla, V., Islam, N., Kaitheri, S.: Developing a mental health index using a machine learning approach: assessing the impact of mobility and lockdown during the COVID-19 pandemic. *Technol. Forecast Soc. Change* **178**, 121560 (2022). <https://doi.org/10.1016/j.techfore.2022.121560>. Epub 2022 Feb 14
- Nuseir, M.T., Aljumah, A., Alshurideh, M.T.: How the business intelligence in the new startup performance in UAE during COVID-19: the mediating role of innovativeness. In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence*. SSDC, vol. 334, pp. 63–79. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67151-8_4n
- Sahu, P.: Closure of Universities Due to Coronavirus Disease 2019 (COVID-19): impact on education and mental health of students and academic staff. *Cureus* **12**(4), e7541 (2020). <https://doi.org/10.7759/cureus.7541>
- Sahut, J.M., Lissillour, R.: The adoption of remote work platforms after the Covid-19 lockdown: new approach, new evidence. *J. Bus. Res.* **154**, 113345 (2023). <https://doi.org/10.1016/j.jbusres.2022.113345>
- Saratchandra, M., Shrestha, A.: The role of cloud computing in knowledge management for small and medium enterprises: a systematic literature review. *J. Knowl. Manag.* **26**(10), 2668–2698 (2022)
- Shah, S. F., Alshurideh, M. T., Al-Dmour, A., Al-Dmour, R.: Understanding the influences of cognitive biases on financial decision making during normal and COVID-19 pandemic situation in the United Arab Emirates. In: Alshurideh, M.T., Hassanien, A.E., Masa'deh, R. (eds.) *The Effect of Coronavirus Disease (COVID-19) on Business Intelligence*. SSDC, vol. 334, pp. 257–274. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-67151-8_15
- Abdallah, S., et al. : Effects of the COVID-19 lockdown and recovery on People's mobility and air quality in the United Arab Emirates using satellite and ground observations. *Remote Sens. Appl. Soc. Environ.* **26**, 100757 (2022). <https://doi.org/10.1016/j.rsase.2022.100757>
- Sindiani, A.M., et al.: Distance education during the COVID-19 outbreak: a cross-sectional study among medical students in North of Jordan. *Ann. Med. Surg. (Lond)*. **59**, 186–194 (2020). <https://doi.org/10.1016/j.amsu.2020.09.036>. Epub 2020 Oct 2
- Singh, M.: Quantifying COVID-19 enforced global changes in atmospheric pollutants using cloud computing based remote sensing. *Remote Sens. App. Soc. Environ.* **22**, 100489 (2021)
- Sultan, R.A., Alqallaf, A.K., Alzarooni, S.A., Alrahma, N.H., AlAli, M.A., Alshurideh, M.T.: How students influence faculty satisfaction with online courses and do the age of faculty matter. In: Hassanien, A.E., et al. (eds.) *AICV 2021. AISC*, vol. 1377, pp. 823–837. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_72
- Taryam, M., et al.: Effectiveness of not quarantining passengers after having a negative COVID-19 PCR test at arrival to Dubai airports. *System. Rev. Pharm.* **11**(11), 1384–1395 (2020)
- Uwizeyimana, D.E.: Analysing the importance of e-government in times of disruption: the case of public education in Rwanda during Covid-19 lockdown. *Eval. Program Plann.* **91**, 102064 (2022)
- Vahdat, S.: The role of IT-based technologies on the management of human resources in the COVID-19 era. *Kybernetes* **51**(6), 20 65–2088 (2022)
- Wong, Y.J., et al.: Spatiotemporal impact of COVID-19 on Taiwan air quality in the absence of a lockdown: Influence of urban public transportation use and meteorological conditions. *J. Clean. Product.* **365**, 132893 (2022)
- World health organization (WHO) (2022). <https://covid19.who.int>

Metaheuristic Algorithms-Based Applications



Metaheuristic Algorithms Based Server Consolidation for Tasks Scheduling in Cloud Computing Environment

Hind Mikram¹, Said El Kafhali¹(✉) , and Youssef Saadi²

¹ Faculty of Sciences and Techniques, Computer, Networks, Modeling, and Mobility Laboratory (IR2M), Hassan First University of Settat, B.P. 539, 26000 Settat, Morocco

{h.mikram,said.elkafhali}@uhp.ac.ma

² Faculty of Sciences and Techniques, Data Science for Sustainable Earth Laboratory, Sultan Moulay Slimane University, Beni Mellal, Morocco
y.saadi@usms.ma

Abstract. Due to the large migration of applications from traditional computing environments to cloud systems, task scheduling has become a well-known research topic. Further, assigning tasks is an essential step in the server consolidation that enables clients' needs to be mapped to suitable resources, such as Physical Machines (PMs) or Virtual Machines (VMs). The difficulty lies in using the available shared computing resources for various tasks without degrading the quality of service (QoS) or increasing the carbon footprint. In addition, the cloud provider needs to use a good task strategy based on the requirements of the client and the IT resources capacity. The best resource utilization and quickest task execution completion times are essential metrics for a cloud provider to maintain the optimum performance of the cloud system. The appropriate solution distributes tasks to the available virtual machine while applying optimization techniques such as the artificial bee colony (ABC), the cuckoo search (CS) algorithm, and particle swarm optimization (PSO). This paper will analyze the algorithms mentioned above based on energy consumption and resource utilization.

Keywords: Server consolidation · Cloud computing · Resource efficiency · Metaheuristic algorithms · Energy consumption · Performance parameters · CloudSim

1 Introduction

Cloud computing has dominated the information technology sector by offering Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). In addition, cloud users can proactively share information from a common database across its deployment models [1]. Cloud data centers require a lot of electricity to operate and provide uninterrupted services

to consumers. According to a survey by International Energy Outlook, energy consumption worldwide is expected to increase by 56 % between 2010 and 2040, also found that the IT industries serve as the primary consumer. Currently, infrastructure-based, hardware-based, and software-based solutions are used to address the issue of rising energy consumption and achieve energy efficiency [2]. It was discovered that energy consumption did not significantly decrease despite the improvement of the hardware and infrastructure [3]. While software-based solutions, including consolidation strategies, maximize resource utilization and minimize the consumed energy. The consolidation techniques allow to turn off of unusable servers and move VMs from underutilized or overutilized servers.

To enhance the performance of resources and maintain the stability of the environment, efficient scheduling among VMs becomes a crucial responsibility [4]. In which load balancing, task scheduling, and resource management fall within the placement step of the consolidation process. Moreover, each user application consists of a collection of tasks that must be scheduled on the appropriate virtual machines [5]. As a result, tasks must be intelligently scheduled while effectively managing the system's processing resources to decrease the completion times and task waiting. However, a poor task assignment approach might make some virtual machines overloaded while others idle, which increases task completion and waiting times. Therefore, heuristic or approximation, meta-heuristic, and machine learning techniques are required to effectively handle the issues.

This work will concentrate on how to solve the load balance and task scheduling issues using metaheuristic algorithms. These techniques have become extremely popular due to their ability to quickly and effectively handle massive, complicated problems. Numerous metaheuristics algorithms have been developed such as ABC, PSO, and CS. However, their inefficiency to detect appropriate information concerning tasks and resources could result in the incorrect assignment of tasks to VMs [6].

Due to the variation of electricity rates across the geographical distribution of the data centers according to the various electricity suppliers and periods, the reduction of energy consumption is a crucial challenge in task scheduling. Furthermore, the management of energy consumption and power pricing is necessary to reduce overall costs. Also, due to the importance of energy consumption in the consolidation process in cloud computing, we will choose it as a parameter of simulation [7]. This study compares metaheuristic methods by evaluating scheduling types across energy consumption and resource utilization metrics on the Cloudsim toolkit.

This paper will divide into the following five sections: The work on consolidation methods and comparative analysis done by other researchers will be reviewed in Sect. 2. The comparative methods will be described in Sect. 3. The energy efficiency will be presented in Sect. 4. The results will be discussed in Sect. 5, and the conclusion of our work will be presented in Sect. 6.

2 Related Work

Due to the varied nature of the cloud resources, task scheduling is one of the crucial issues in the server consolidation process. Numerous methods have been proposed to handle the task scheduling problem in the cloud computing environment. Each scheduling strategy has advantages and disadvantages due to the QoS criteria considered. This article will discuss a few recent studies of scheduling problems in a cloud system. The original CRO (Chemical Reaction Optimization) and traditional PSO methods have been improved [8], which suggests a scheduling technique named CR-PSO whose purpose of optimizing the makespan, average execution time, cost, and energy factors. The algorithm does not consider the task dependency. Additionally, they do not evaluate the performance of energy and power consumption. The authors [9] suggested a task scheduling method based on the Harris Hawks optimization technique. This algorithm simultaneously increases the QoS and the level of service agreement. They consider energy consumption and other QoS factors to compare the suggested approach with other algorithms like PSO and TLBO (Teaching-Learning Based Optimization).

In Optim [10], a hybrid Fuzzy Particle Swarm Optimization Genetic Algorithm (FPSO-GA) was proposed. The goal of [11] is to identify the best method for a multi-criteria scheduling problem. The authors consider energy consumption and makespan simultaneously to solve cloud workflow scheduling. The study in [12] suggests GALCS, a GA mixed with LCS (Longest Common Subsequence) selection to increase search effectiveness. They create a basic system for allocating processors depending on the quantity of task data and the processing power of VMs, but they neglect the dependence between tasks. The authors compare the ant colony optimization, the ABC algorithm, the PSO algorithm, the crow search algorithm, the penguin swarm optimization algorithm, and the genetic algorithm. They have determined that the Crow Search algorithm is the most advantageous method in resource utilization cost and makespan parameters. In [13], the authors examine and experimentally evaluate some of the most recent scheduling mechanisms in cloud computing concerning throughput (the number of completed tasks per unit of time) and Makespan (the time difference between the beginning and end of a collection of tasks). They enhance the comparison by assessing the Average Resource Utilization Ratio (ARUR) to the methods under evaluation without considering energy consumption. They also identify and recommend parameters that can improve resource usage and increase cloud service providers' ability to generate revenue.

The proposed technique in [14] tends to allocate tasks to the VM with the lowest execution cost within a deadline limitation and then deploy the chosen VM on the busiest PM within a capacity constraint. The evaluation parameters are resource utilization, degree of imbalance, makespan, and execution cost. This algorithm does not test energy consumption, reliability, and security [15]. The MOTS-ACO algorithm has been used in [16] to offer a new multi-objective task scheduling algorithm based on an improved ant colony optimizer in the cloud data center. The objectives of this algorithm are to improve energy consumption,

load balancing, makespan, and turnaround time in cloud data centers, identify the best approach to the task scheduling issue, and look for the best trade-off efficiency approach that achieves a balance between competing goals. However, when it comes to loading balancing with huge tasks, the proposed algorithm does not perform effectively.

It can be necessary to evaluate numerous approaches to find the best method in terms of resource utilization and energy consumption. In this paper, the choice of the compared algorithms was based on the robustness and simplicity of those methods. For instance, CS has quick convergence and speed, which means its robustness in global optimization [19]. ABC handle a wide range of situations and issue types. Frequently finds effective solutions even when the search space is quite complicated [26]. PSO offers paths that are more stable in comparison to other traditional techniques [27].

3 Metaheuristic Algorithms Backgrounds

3.1 Artificial Bee Colony Algorithm

The original developers of this algorithm are Karaboga and Basturk. The employed, onlooker, and scout bees work together to search for the optimal food source by replicating honey bees' intelligent foraging behavior. The vector d_i with ND dimensional design variables is the i th solution, starting with a random population size N [17]. The following formulas are used to update the population after each iteration:

$$d_{ij} = d_{ij} + r_{ij} \cdot (d_{ij} - d_{lj}) \quad (1)$$

The random indexes represented by $l \in \{1, 2, \dots, BN\}$ and $j \in \{1, 2, \dots, DN\}$. r_{ij} , a random number that falls between $[-1, 1]$, is used to regulate food in d_{ij} . The fitness of two locations for visible food is compared using the constant k .

The employed bee number is BN , and D stands for the optimization parameter number. The location d_{ij} is governed by d_{ij} and d_{kj} , as can be observed from Eq. 1.

3.2 Particle Swarm Optimization Algorithm

PSO was established in 1995 by electrical engineer Russell Eberhart and social psychologist James Kennedy. It has seen numerous improvements since its beginning. This search produced a random set of possible solutions. The term "swarm" refers to this group of potential solutions, whereas the term "particle" refers to each potential solution.

Two different types of learning have an impact on the PSO search. Each particle gets information from other particles as well as from its own mobility experiences. While learning through one personal experience is considered cognitive learning, learning from others is referred to as social learning. In cognitive learning, the particle saves the best solution it has found on its own experience

thus far, known as pbest. The best solution for every particle in the swarm is stored in gbest as a result of social learning. Any particle changes its direction and magnitude by velocity. This is the rate at which the position is changing with time. Time is the iteration in terms of the PSO [18].

3.3 Cuckoo Search Algorithm

The creator of the Cuckoo Search Algorithm is XinShe Yang and Suas Deb. CS has considerably demonstrated itself in global optimization due to its quick convergence and speed. CS has been used to address many engineering issues, including scheduling. The main goal of this algorithm is to lay the cuckoo eggs in another cuckoo's nest to boost the chances that the eggs will hatch. Three guiding concepts control the original Cuckoo Search:

- Each cuckoo laid an egg, which then was deposited in a random nest.
- The next generation or solution will be carried by the best nest.
- The number of host nests is fixed and there is a chance that a host will discover strange eggs (0 or 1). In this situation, the host can either throw a cuckoo egg or depart the nest and create a new one.

The CS algorithm uses three parameters:

- $P_\alpha \in [0, 1]$ probability that the worst nest will be abandoned.
- $\alpha > 0$ step size, which is often taken at $\alpha > 1$, depends on the magnitude of the problem.
- λ is a random step length.

In the CS algorithm, the probability of changing and finding defines the search locally or globally. The value of $P_\alpha = 0.25$ was typically used for local searches, although a global search can be investigated more effectively for $1 - P_\alpha$ time, i.e. (0.75). The global search is assisted by the levy flight function, which made numerous multi-objective studies have demonstrated good global convergence [19].

4 Energy Efficiency Model

The need for cloud data centers has increased, resulting in the expansion of web-based software applications and the internet. Numerous servers in cloud data centers were available to consumers, which confirms the vast amount of energy required to run them. However, server utilization varies with time, making energy management a crucial task for managing cloud resources from the point of financial viability [20].

The energy used by IT equipment, such as PMs and network equipment, and the energy used by non-IT equipment together make up a data center's total energy usage, like cooling systems. The systems that manage lightning, fire, and

electrical equipment, are said to use only a little power [21]. Furthermore, workload influences the energy efficiency of multi-core systems. A core will operate at its highest frequency and voltage when it is under its maximum workload, resulting in significant energy consumption. The task is split among the processors, resulting in a faster processing time with less power usage. The only solution for this issue in multi-core systems is to equally distribute the workload while maintaining the ideal CPU frequency with the lowest energy consumption ratio. An idle server uses two-thirds as much energy as one that is fully utilized and under load. It is interesting that different power models for physical servers cause varying degrees of idle power and dynamic power consumption use. By consolidating virtual machines, it is possible to reduce energy consumption without sacrificing resource availability, risking the provider's credibility. High voltage usage makes servers hot and causes them to last less time. To reduce active and idle server power consumption, resource usage should be optimized based on the processing capabilities of the servers [20].

4.1 Energy Consumption

The following is the power consumption equation:

$$P_i(s) = P_{idle}(s) + \mu.(P_{max}(s) - P_{idle}(s)) \quad (2)$$

where s is the current CPU utilization and $1 - P_{max}$ is the maximum power that a server can use while operating at 100% CPU utilization. We define CPU utilization as a function of time because it varies over time [22]. As a result, Eq. 3 establishes how much energy a physical machine (such as a server) uses overall:

$$P_{server} = \sum_{i=1}^n P(s(t)) \quad (3)$$

where n indicates the number of servers and $P(s(t))$ is the energy used by a specific PM i at utilization s .

4.2 Resource Utilization

Resource utilization is a parameter of performance used to evaluate how well resources are being used. A high resource usage rate indicates that a cloud provider can make the most revenue. Equation 4 determines the resource utilization (RU) [23].

$$RU = \frac{\sum_{i=1}^{nbrofVM} CT_i}{makespan \times nmbrofVM} \quad (4)$$

With

$$CT_i = \frac{\sum_{j=1}^{nbroftasks} Task_j.length}{vm_i.pesnumber \times vm_i.mips} \quad (5)$$

and

$$makespan = \max_{1 \leq i \leq nbrofVM} \{CT_i\} \quad (6)$$

5 Simulations Results and Analysis

5.1 Simulation Environment

Task scheduling significantly depends on the size of the tasks as well as their numbers. However, server configuration also has a significant impact on how effectively the resources are used. It specifies the types and number of VMs that can be run simultaneously on the server at any given time to carry out the activities [24].

The CloudSim simulator was employed to compare the performance of the chosen techniques. The running system has an Intel Core i5- 6300U processor, clocking at 2.4 GHz, and 20 Go of RAM. The task length considered varies between 1000 and 2000; With 100, 300, and 700 as the number of these tasks; The number of heterogeneous VMs is fixed at 10, split equally among 2 PMs under resources constraint. One VM type was taken in which the MIPS varies from 100 to 1000. In this experiment, the task shared the space, and the VM shared the time.

5.2 Results and Discussion

The scheduling algorithms' effectiveness is based on various parameters such as the level of SLA violation, energy consumption, makespan, throughput, and other factors [25]. This paper chooses energy consumption and resource utilization as performance metrics. Following, the evaluation and discussion of the outcomes of the chosen strategies.

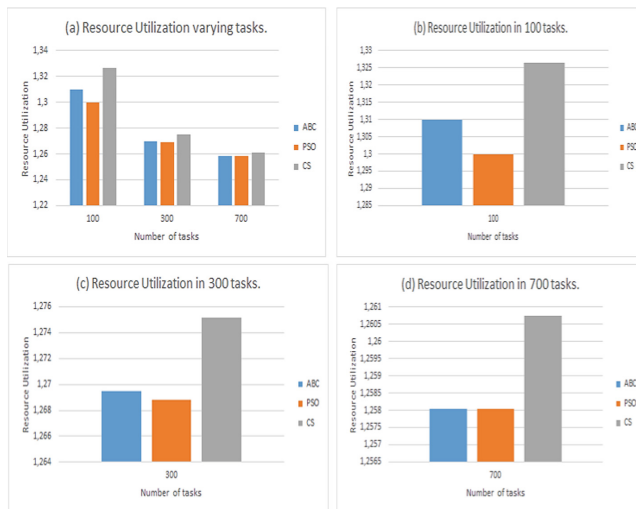


Fig. 1. Resource utilization varying task scheduling

From Fig. 1 (a), the resource utilization of the compared algorithms decreases as the number of tasks increases. In Figs. 1 (b) and (c), the results show that the resource utilization of CS is much better than that of ABC and PSO. In Fig. 1 (d), PSO and ABC obtain less resource utilization than CS. The resource utilization parameter should have the maximum value to avoid resource wastage. Therefore, the CS algorithm has the optimum utilization of resources, whereas PSO has the worst resource usage.

The energy consumption of all the examined scheduling methods is shown in Fig. 2 (a). Where the energy consumed increased, when the number of tasks augment.

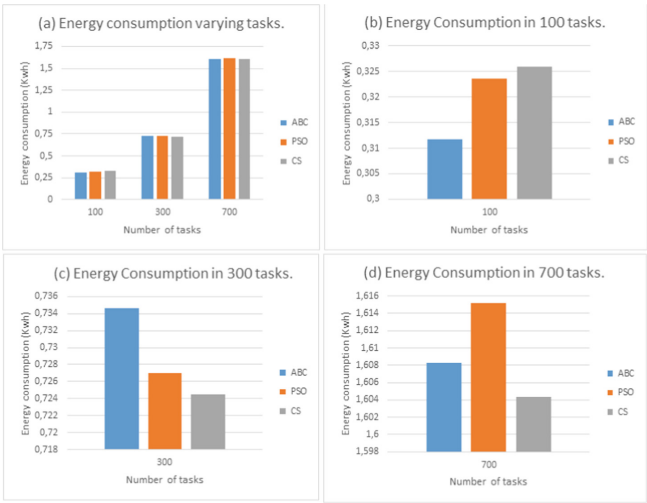


Fig. 2. Energy consumption varying task scheduling

Figure 2 (b) shows the value of the energy consumed for the three meta-heuristic methods in 100 tasks. The ABC minimizes energy consumption better than PSO and CS, respectively. In contrast, in Fig. 2 (c), the CS decreases the energy consumption value less than ABC and PSO. The highest value for PSO was in Fig. 2 (d), where the CS had a small value of the consumed energy.

The best algorithm in these scenarios was CS because it had a small value. This means reducing the wastage of the consumed energy. Therefore, the CS had optimal utilization in 300 and 700 tasks, in other words, in medium and large workloads. Whereas ABC is the best in the small number of tasks (100 tasks), and PSO is the worst in the high number of tasks. Through the discussion above, the following points are underlined in Table 1.

Where $\checkmark$ indicates that the algorithm is good. X determines the worst algorithm. $\sim$ defines in the middle of workload types.

Table 1. Energy consumption over workloads

Algorithms	Small workload	Medium workload	Large workload
ABC	✓	X	~
PSO	~	✓	X
CS	X	✓	✓

6 Conclusion and Future Work

This paper compared three metaheuristic methods used for task scheduling by employing CloudSim as a simulator. The techniques used for the analysis named PSO, ABC, and CS. The workload employed for the studies was set up with various levels of task and VM heterogeneity. Among the compared scheduling approaches, the CS algorithm outperformed as compared to the rest of the scheduling methods. In addition, energy consumption increased as the number of tasks increased, in contrast to resource utilization, which decreased. In our future work, we plan to suggest a hybrid task scheduling technique that takes more objectives and compares it to the most recent approaches.

References

1. Mikram, H., El Kafhali, S., Saadi, Y.: Server consolidation algorithms for cloud computing: taxonomies and systematic analysis of literature. *Int. J. Cloud Appl. Comput. (IJCAC)* **12**(1), 1–24 (2022)
2. El Kafhali, S., Salah, K.: Modeling and analysis of performance and energy consumption in cloud data centers. *Arab. J. Sci. Eng.* **43**(12), 7789–7802 (2018)
3. El Kafhali, S., Salah, K.: Performance analysis of multi-core VMs hosting cloud SaaS applications. *Comput. Stand. Interfaces* **55**, 126–135 (2018)
4. Hanini, M., El Kafhali, S.: Cloud computing performance evaluation under dynamic resource utilization and traffic control. In: *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications*, pp. 1–6 (2017)
5. El Kafhali, S., El Mir, I., Salah, K., Hanini, M.: Dynamic scalability model for containerized cloud services. *Arab. J. Sci. Eng.* **45**(12), 10693–10708 (2020)
6. Adhikari, M., Amgoth, T.: Heuristic-based load-balancing algorithm for IaaS cloud. *Future Gener. Comput. Syst.* **81**, 156–165 (2018)
7. Hussain, M., Wei, L.F., Rehman, A., Abbas, F., Hussain, A., Ali, M.: Deadline-constrained energy-aware workflow scheduling in geographically distributed cloud data centers. *Future Gener. Comput. Syst.* **132**, 211–222 (2022)
8. Dubey, K., Sharma, S.C.: A novel multi-objective CR-PSO task scheduling algorithm with deadline constraint in cloud computing. *Sustain. Comput. Inform. Syst.* **32**, 100605 (2021)
9. AL-Amodi, S., Patra, S.S., Bhattacharya, S., Mohanty, J.R., Kumar, V., Barik, R.K.: Meta-heuristic algorithm for energy-efficient task scheduling in fog computing. In: Dhawan, A., Tripathi, V.S., Arya, K.V., Naik, K. (eds.) *Recent Trends in Electronics and Communication*. LNEE, vol. 777, pp. 915–925. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-2761-3_80

10. Mirmohseni, S.M., Tang, C., Javadpour, A.: FPSO-GA: a fuzzy metaheuristic load balancing algorithm to reduce energy consumption in cloud networks. *Wireless Pers. Commun.* **127**, 2799–2821 (2022). <https://doi.org/10.1007/s11277-022-09897-3>
11. Singh, H., Tyagi, S., Kumar, P., Gill, S.S., Buyya, R.: Metaheuristics for scheduling of heterogeneous tasks in cloud computing environments: analysis, performance evaluation, and future directions. *Simul. Model. Pract. Theory* **111**, 102353 (2021)
12. Xia, X., Qiu, H., Xu, X., Zhang, Y.: Multi-objective workflow scheduling based on genetic algorithm in cloud environment. *Inf. Sci.* **606**, 38–59 (2022)
13. Ibrahim, M., et al.: A comparative analysis of task scheduling approaches in cloud computing. In 2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID), pp. 681–684. IEEE (2020)
14. Alboaneen, D., Tianfield, H., Zhang, Y., Pranggono, B.: A metaheuristic method for joint task scheduling and virtual machine placement in cloud data centers. *Future Gener. Comput. Syst.* **115**, 201–212 (2021)
15. Salah, K., El Kafhali, S.: Performance modeling and analysis of hypoexponential network servers. *Telecommun. Syst.* **65**(4), 717–728 (2017)
16. Elsedimy, E., Algarni, F.: MOTS-ACO: an improved ant colony optimiser for multi-objective task scheduling optimisation problem in cloud data centres. *IET Netw.* **11**(2), 43–57 (2022)
17. Meng, Z., Li, G., Wang, X., Sait, S.M., Yıldız, A.R.: A comparative study of metaheuristic algorithms for reliability-based design optimization problems. *Arch. Comput. Methods Eng.* **28**(3), 1853–1869 (2021)
18. Bansal, J.C., Singh, P.K., Pal, N.R. (eds.): *Evolutionary and Swarm Intelligence Algorithms*. SCI, vol. 779. Springer, Cham (2019). <https://doi.org/10.1007/978-3-319-91341-4>
19. Kumar, M., Suman.: Hybrid Cuckoo Search Algorithm for Scheduling in Cloud Computing. *CMC-Comput. Mater. continua* **71**(1), 1641–1660 (2022)
20. Renugadevi, T., Geetha, K., Prabakaran, N., Siano, P.: Carbon-efficient virtual machine placement based on dynamic voltage frequency scaling in Geo-Distributed cloud data centers. *Appl. Sci.* **10**(8), 2701 (2020)
21. Feng, H., Deng, Y., Li, J.: A global-energy-aware virtual machine placement strategy for cloud data centers. *J. Syst. Archit.* **116**, 102048 (2021)
22. Saadi, Y., El Kafhali, S.: Energy-efficient strategy for virtual machine consolidation in cloud environment. *Soft Comput.* **24**(19), 14845–14859 (2020)
23. Mapetu, J.P.B., Chen, Z., Kong, L.: Low-time complexity and low-cost binary particle swarm optimization algorithm for task scheduling and load balancing in cloud computing. *Appl. Intell.* **49**(9), 3308–3330 (2019)
24. Adhikari, M., Amgoth, T.: Heuristic-based load-balancing algorithm for IaaS cloud. *Future Generat. Comput. Syst.* **81**, 156–165 (2018)
25. Ibrahim, M., et al.: A comparative analysis of task scheduling approaches in cloud computing. In 2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID), pp. 681–684. IEEE (2020)
26. Karaboga, D., Akay, B.: A comparative study of artificial bee colony algorithm. *Appl. Math. Comput.* **214**(1), 108–132 (2009)
27. Goyal, S., et al.: An optimized framework for energy-resource allocation in a cloud environment based on the whale optimization algorithm. *Sensors* **21**(5), 1583 (2021)



Spectrum Recovery Improvement in Cognitive Radio Using Grey Wolf Optimizer

Gehan Gamal¹ (✉), Mohamed F. Abdelkader², Abdelazeem A. Abdelsalam¹,
and Ahmed Magdy¹

¹ Electrical Engineering Department, Faculty of Engineering, Suez Canal University, Ismailia,
Egypt

gehangamal@eng.suez.edu.eg

² Electrical Engineering Department, Faculty of Engineering, Port Said University, Port Said,
Egypt

Abstract. By using spectrum gaps caused by the underuse of frequency spectrum, cognitive radio is one of the most promising methods to enable opportunistic spectrum access for secondary users. In order to prevent interference with primary users, a cognitive radio unit must frequently sense the spectrum. Wideband spectrum sensing techniques that use compression are gaining popularity because they eliminate the need for high-rate analog-to-digital converters, which in turn reduces the complexity and power consumption of the cognitive radio. The Grey Wolf Optimizer is one of the most recent optimization techniques, that is capable of resolving the compressive spectrum recovery technique's optimization issue. By comparing the effectiveness of algorithms for detecting the spectrum, the simulation results in this study have shown that the grey wolf optimizer algorithm performed better than basis pursuit denoising approach.

Keywords: Cognitive radio · Wideband spectrum sensing · Compression · Basis pursuit denoising approach · Grey Wolf Optimizer

1 Introduction

Because primary users (PUs) or licensed users do not always use their channels, the research [1] revealed that the frequency spectrum has been underutilized. Through gaps in the spectrum that primary users leave open, secondary users (SUs) or unlicensed users may be able to communicate [2]. As a means of addressing the issue of global spectrum scarcity through opportunistic spectrum access, cognitive radios (CR) have drawn more and more attention. A CR unit must frequently detect spectral holes by sensing the spectrum that can be used for transmission when PU is not present, so it must understand its surroundings [3]. By scanning a wider range of frequency bands, wideband spectrum sensing improves the SU's chances of finding unoccupied bands. However, the Nyquist theorem states that it necessitates high-speed sampling of the spectrum [4]. As a result, the required high-speed analogue to digital converter (ADC) is subjected to hardware

technological limitations, increasing the overhead in terms of energy and communication. Several application sectors, including wireless communication, have effectively used compressed sensing (CS) [5] and sparse representation techniques to lower data acquisition costs [6]. It enables sub-Nyquist sampling rates of the spectrum using an analogue to information converter (AIC), while employing CS reconstruction methods to effectively recreate the original spectrum by taking advantage of the sparsity brought on by underutilization [7]. A few non-zero coefficients in sparse signals hold the majority of the signal energy. Additionally, for the purpose of applying the CS theory [8], the signal itself need only be compressible and sparsely represented in a specified transform domain [9]. The initial CS methods only consider the signal's sparsity as prior information about the unknown signal [5]. In order to further decrease the necessary sample rate, several methods were developed to improve the compressive spectrum sensing's efficacy by utilizing more redundancy. In cooperative networks, this also includes utilizing spatial correlation [10, 11]. The minimization problem of the spectrum recovery process could be solved by solving Basis Pursuit (BP) optimization problem [12]. If an environment of noisy observation is assumed, this problem is transformed to the Basis Pursuit Denoising (BPDN) optimization problem [13]. Numerous optimization methods have been used to wireless communication, including Particle Swarm Optimization (PSO) [14] and Gravitational Search Algorithm (GSA) [15]. One of the most current metaheuristic swarm intelligence techniques is the grey wolf optimizer (GWO) [16]. Due to its impressive advantages over other swarm intelligence technique, the fact that it requires no derivation information in the initial search and has a small number of parameters, it has been extensively suited for a variety of optimization tasks. The grey wolves that search for the best route to hunt their preys in nature are the inspiration for the GWO algorithm. The GWO algorithm uses a similar method in nature, where it organizes the various positions in the wolf pack according to the pack hierarchy [17]. According to the several wolf roles that aid in the advancement of the hunting process, the GWO pack is organized into four groups. Alpha, Beta, Delta, and Omega are the four groupings, with Alpha standing in as the best hunting solution to date [18].

In this research, to demonstrate the efficacy of this technique, we compare the performance of the GWO algorithm with that of the BPDN strategy for solving the optimization problem of the spectrum recovery process in a CR network.

2 Proposed Methodology

By taking into account a CR network in which each SU terminal checks U non-overlapping channels across a broadband locally. Finding out if any of these channels are being used or are available for opportunistic usage is the challenge of spectrum sensing.

Take into account M active PUs, the following model can be used to describe the signal that all PUs sent to the CR [19]:

$$x(t) = \tilde{x}(t) + w(t) \quad (1)$$

where $\tilde{x}(t)$ is the PUs' noise-free received signal and $w(t)$ is the CR's additive white Gaussian noise (AWGN).

Given (1)'s Discrete Fourier Transform (DFT), the sensed signal can be shown as follows.

$$X = \sum_{m=1}^M \tilde{H}_m \tilde{S}_m + W \quad (2)$$

where $\tilde{H}_m$ is $U \times U$ diagonal matrix, whose principal diagonal is the DFT at U points of $\tilde{h}_m$. X , $\tilde{S}_m$ and W are, respectively, the DFT transformations of x , $\tilde{s}_m$ and w .

Equation (2) may be layered to create a matrix as

$$X = \tilde{H} \tilde{S} + W \quad (3)$$

where $\tilde{H}$ is a $U \times UM$ matrix and $\tilde{S}$ is a $UM \times 1$ vector. We presume that each channel has no more than one operational PU transmitter.

It is now possible to depict the sensed spectrum as

$$X = HS + W \quad (4)$$

Compressive spectrum sensing uses a CR receiver to gather compressed linear combination measurements of the signal samples x . From the signal X that was received, the spectrum vector S can be estimated.

In the following, CR gathers $L \times 1$ time samples of the measurement vector y from the input x , where $M < L \ll U$

$$y = \varphi x \quad (5)$$

where φ is a $L \times U$ random measurement matrix has elements that can be independently distributed random variables with the same distribution. φ must adhere to the restricted isometry property (RIP) in order to guarantee flawless recovery of the sparse signal, which stipulates that if an isometry constant δ_s with a value of $0 < \delta_s < 1$ exists, it must meet the following inequality [20]

$$(1 - \delta_s) \|x\|_2^2 \leq \|\varphi x\|_2^2 \leq (1 + \delta_s) \|x\|_2^2 \quad (6)$$

where $\|\cdot\|_2$ denotes the l_2 -norm.

To do this, choose a Gaussian measurement matrix, and ensure that approximately $C L \log(U/M)$ measurements have been made, where C is a constant.

There could be an endless number of solutions to this equation of the system because there are fewer rows L than columns U in the sensing matrix, but we are only concerned with the sparsest one by using l_0 -norm minimization problem, which is an NP-hard issue. The following convex optimization issue, like in [5], can be solved to relax the problem to a l_1 -norm minimization that

$$\min_x \|x\|_{l_1} \text{ subject to } y = \varphi x \quad (7)$$

Numerous convex relaxation methods, including Basis Pursuit (BP) [12–21], can address this convex optimization problem [22].

This issue is changed to a Lagrange form known as the BP denoising under the assumption that the observing environment is noisy (BPDN) [13]

$$\min_x \|x\|_{l_1} + \lambda \|y - \varphi x\|_{l_2}^2 \tag{8}$$

where the penalty parameter λ can be utilized to renegotiate between minimizing the L_2 norm term and ensuring the spectrum’s sparsity, if the variance of the noise is known, as was demonstrated in [21], λ might be approximated. The Least Absolute Shrinkage and Selection Operator (LASSO) and the BPDN optimization issue in (8) are extremely similar, so we can utilize this for solving our optimization problem.

3 Optimization Process Using GWO Algorithm

The formation of a random grey wolf population is the first step in the GWO optimization process (candidate solutions). The wolves recognize the likely location of the prey across the iterations (optimum solution). The distance between the grey wolves and their prey determines how they should position themselves [22].

GWO, a brand-new meta-heuristic technique for optimization introduced by Mirjalili, can be used to address the optimization problem. Four different breeds of grey wolves including alpha, beta, delta, and omega are used to model the framework of the leadership. Furthermore, the three fundamental elements of hunting are looking for prey, surrounding prey and attacking prey are utilized. The social structure of grey wolves is shown in Fig. 1.

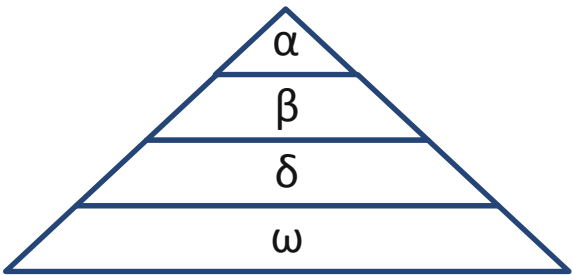


Fig. 1. The social structure of grey wolves.

3.1 Modeling the GWO Algorithm Mathematically

- **Social Structure**

The best answer, alpha (α), is used to quantitatively simulate the social hierarchy of grey wolves in the GWO algorithm. Beta (β) and delta (δ) are, respectively, regarded the second and third best solutions. And another solution, which is seen as being omega (ω). The GWO algorithm uses α , β , and δ to lead hunting (optimization), and ω wolves follow them.

• Embedding the Prey

Grey wolves, as noted, surround their victim while hunting. The grey wolves' circling movement when searching for prey can be expressed as [17]

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right| \quad (9)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (10)$$

where t = repetition count, $\vec{A}$ and $\vec{C}$ = vectors of coefficients, $\vec{X}_p$ = position of the prey in a vector, $\vec{X}$ = positional vector for the grey wolves; and $\vec{D}$ = a calculated vector was used to specify the grey wolf's new location. The vectors of coefficients could be calculated as

$$\vec{A} = 2 \vec{b} \cdot \vec{r}_1 - \vec{b} \quad (11)$$

$$\vec{C} = 2 \cdot \vec{r}_2 \quad (12)$$

where $\vec{b}$ = vector configured to drop linearly from 2 to 0 across the iterations, and $\vec{r}_1$ and $\vec{r}_2$ = random vectors in [0,1]. We use the assumption that candidates α (best candidate for the solution), β , and δ have more information about the likely location of the prey in order to imitate the hunting behavior of grey wolves. To attain the optimal position in the decision space, the algorithm forces others (such as omega wolves) to update their locations while saving the three best solutions so far [23]. This kind of hunting behavior can be described by Mirjalili in the optimization algorithm [16].

$$\vec{D}_\alpha = \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right|, \vec{D}_\beta = \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right| \text{ and } \vec{D}_\delta = \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right| \quad (13)$$

$$\vec{X}_1 = \vec{X}_\alpha - A_1 \cdot (\vec{D}_\alpha), \vec{X}_2 = \vec{X}_\beta - A_2 \cdot (\vec{D}_\beta) \text{ and } \vec{X}_3 = \vec{X}_\delta - A_2 \cdot (\vec{D}_\delta) \quad (14)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (15)$$

Based on the locations of α , β , and δ in the decision space, a final position (solution) is defined as being inside a circle. To put it another way, wolves α , β , and δ make estimates

of the locations of prey and other wolves before updating those estimates randomly near the prey.

• Predator Attacking

Grey wolves complete their hunt by assaulting their prey until it stops moving, as was previously indicated. To simulate the attacking procedure, the value of $\vec{b}$ could be decreased in different iterations [24]. During iterations, $\vec{b}$ is reduced from 2 to 0 and $\vec{A}$ is a random value in the range $[-2b, 2b]$. The future position of a search agent can be anywhere between its current position and the position of the prey when the random values of $\vec{A}$ are in the range $[-1, 1]$.

The search agent can update its position using the positions using the GWO method α , β and δ . While it is true that the GWO algorithm's process of surrounding restricts the solutions to those near local optima, GWO also offers a wide range of additional operators that can lead to the discovery of novel solutions. Grey wolves will attack the prey when $|\vec{A}| < 1$.

• The Hunt for the Prey

Grey wolves first divide to investigate the location of the prey, then converge to attack it. To quantitatively simulate the divergence of grey wolves, $\vec{A}$ could be expressed as a random vector which is greater than 1 or less than -1 to have the search agent depart from the target, emphasizing the importance of the global search in GWO. When $|\vec{A}| > 1$, grey wolves are compelled to separate from their prey in an effort to locate a fitter prey.

3.2 GWO Algorithm Optimization Procedure

The GWO optimization procedure begins with the generation of a random population of grey wolves (candidate solutions) [24]. The wolves in iterations α , β , and δ estimate where the prey is likely to be (optimum solution). Based on how far they are from their prey, grey wolves adjust where they are. When $|\vec{A}| > 1$, the potential remedies deviate from the target and when $|\vec{A}| < 1$, The potential solutions all lead to the prey. Some notes on how the GWO method solves optimization problems might be summed up as follows:

- The GWO algorithm uses the idea of social hierarchy to rank the answers and keep the better ones for the current iteration.
- The answer is defined by the encircling process as a 2D neighbor with a circle form.
- Grey wolves are assisted in defining various hyper-spheres with random radii by the random parameters (A and C).
- The GWO algorithm's hunting strategy enables grey wolves to choose the most likely location of the prey (optimum solution).

- The adaptable values of parameters A and b ensure exploration and exploitation of the GWO algorithm and switching between them is made simple.

The GWO algorithm's flowchart containing information on the optimization procedure is shown in Fig. 2.

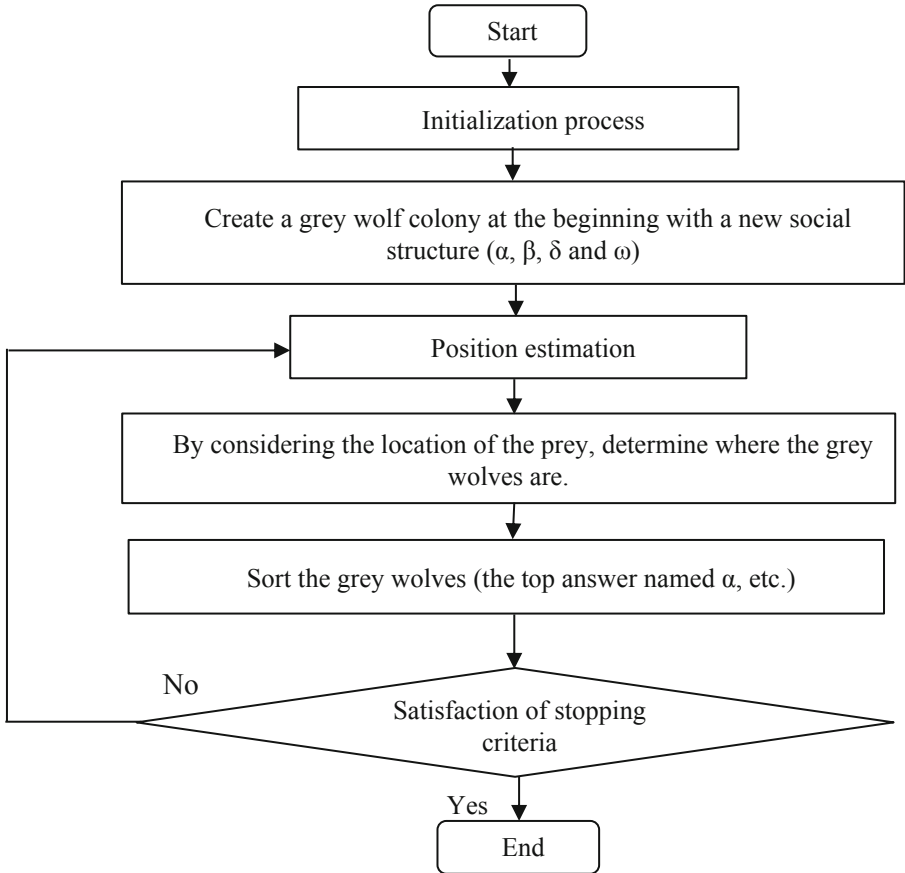


Fig. 2. Algorithm flowchart for GWO

4 Experimental Results

By using numerical simulations, we assess the performance of the GWO method for the spectrum recovery process in comparison to the BPDN approach. In each of our studies, we take into account a spectrum with $U = 128$ sub-channels [19]. Multipath fading channels with $N_p = 3$ taps are used, each tap's gain is derived from a Rayleigh distribution. The received signal is distorted by an additive white Gaussian noise (AWGN). The

measurements number L divided by the signal's dimension U is known as the compression ratio (cr) and the Signal to Noise Ratio (SNR) is defined as the signal power over the whole bandwidth normalized by the noise power.

We see that in our problem, the probability of detection P_d relates to the likelihood of detecting the active PUs, but the probability of false alarm P_{fa} refers to the likelihood of labelling an empty channel as occupied.

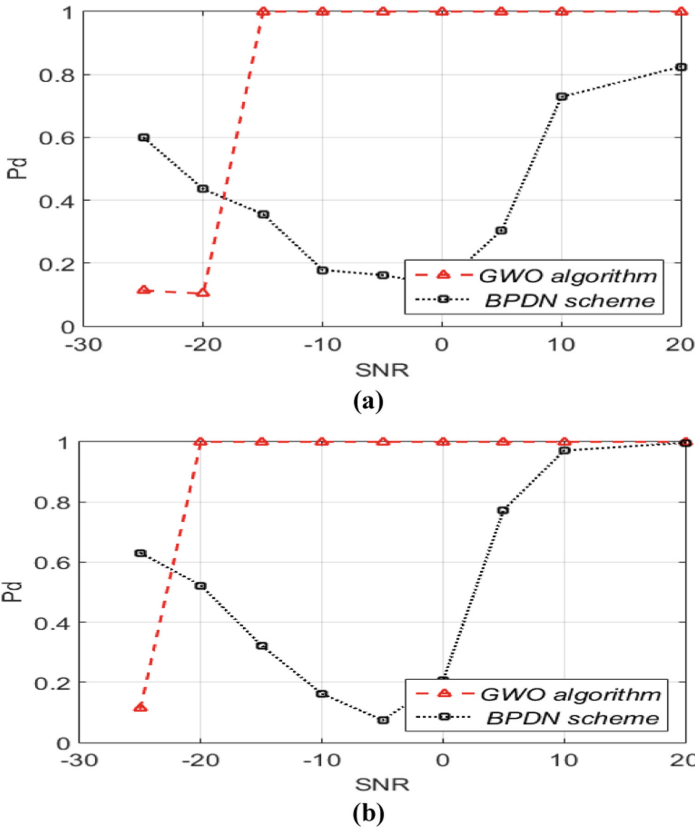


Fig. 3. Probability of detection at various SNR levels in dB at probability of false alarm = 0.1 and at compression ratio of (a) $cr = 0.3$ and (b) $cr = 0.5$.

In our simulation of the GWO algorithm, we take into account a maximum of 500 iterations and 30 search agents. Figure 3 displays the performance of detection accuracy for various SNR values at compression ratios of 0.3 and 0.5. As shown in Fig. 3(a), the detection of the spectrum for the GWO algorithm achieves superior performance 100% at SNR values of -15 and higher, at variance the BPDN scheme reached 80% as the maximum ratio at $SNR = 20$ dB at $cr = 0.3$. And in Fig. 3(b), the probability of detection of the GWO algorithm achieves the maximum value from $SNR = -20$ dB at $cr = 0.5$, in contrast the probability of detection for the BPDN scheme reaches the maximum value only at $SNR = 20$ dB.

In the Fig. 4 the variation in performance of detection at various compression ratio levels the same 0.1 false alarm probability at SNR = -20 dB and 10 dB. And as shown in Fig. 4(a) when SNR = -20 dB, the performance of the GWO algorithm achieves the maximum at $cr = 0.4$ compared to the BPDN scheme which P_d value still around 50%. And in Fig. 4(b) when SNR = 10 dB, the probability of detection achieves superior over all range of the compression ratio in the GWO algorithm compared to the BPDN scheme which P_d reaches the maximum value at $cr = 0.75$ only.

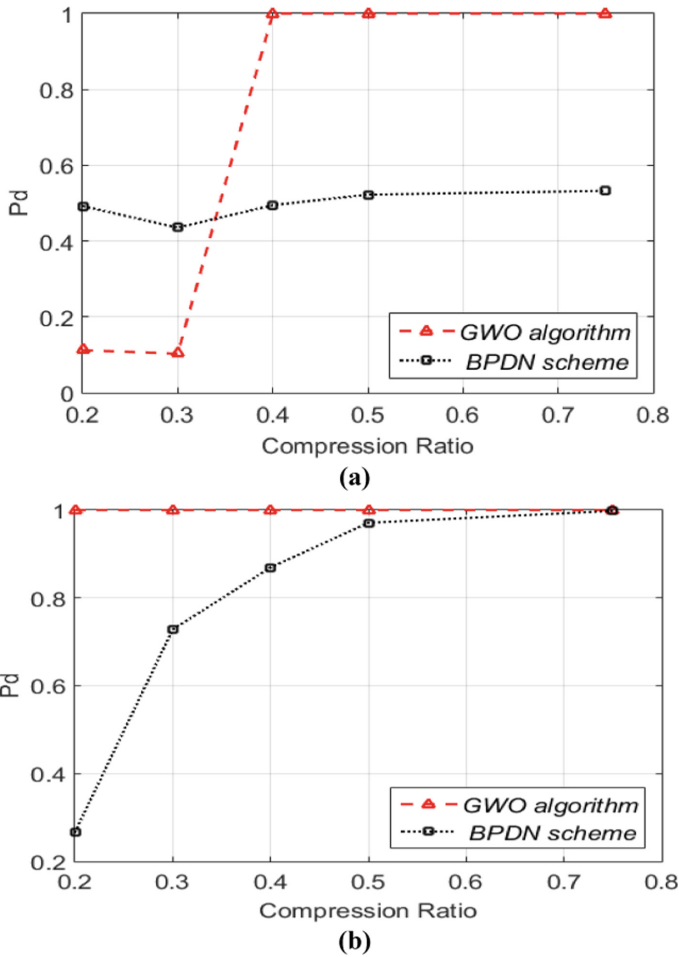


Fig. 4. Probability of detection at various compression ratios at probability of false alarm = 0.1 at (a) SNR = -20 dB (b) SNR = 10 dB.

5 Conclusion

In this study, we solve the minimization problem for the compressed spectrum sensing recovery process using the GWO algorithm. Additionally, it is compared to the BPDN approach to solve the minimizing recovery problem. The simulation results and discussions demonstrated that the GWO algorithm has superior performance and the ability to solve the minimization problem for the spectrum recovery process, that the accuracy ratio reaches 100% at a wide range of SNR from -20 dB to 20 dB and over the full range of compression ratios in comparison to the BPDN algorithm, which only achieve this precision at SNR values of 20 . As can be shown, our suggested strategy performs better than other techniques.

References

1. Akyildiz, I.F., Lee, W.Y., Vuran, M.C., Mohanty, S.: NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey. *Comput. Netw.* **50**(13), 2127–2159 (2006)
2. Wang, B., Liu, K.R.: Advances in cognitive radio networks: a survey. *IEEE J. Sel. Top. Signal Process.* **5**(1), 5–23 (2010)
3. Haykin, S.: Cognitive radio: brain-empowered wireless communications. *IEEE J. Sel. Areas Commun.* **23**(2), 201–220 (2005)
4. Por, E., van Kooten, M., Sarkovic, V.: Nyquist–Shannon sampling theorem. *Leiden University* **1**(1) (2019)
5. Tian, Z., Giannakis, G.B.: Compressed sensing for wideband cognitive radios. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007), vol. 4, pp. IV-1357. IEEE (2007)
6. Qin, Z., Fan, J., Liu, Y., Gao, Y., Li, G.Y.: Sparse representation for wireless communications: a compressive sensing approach. *IEEE Signal Process. Mag.* **35**(3), 40–58 (2018)
7. Kirolos, S., Ragheb, T., Laska, J., Duarte, M.F., Massoud, Y., Baraniuk, R.G.: Practical issues in implementing analog-to-information converters. In: 2006 6th International Workshop on System on Chip for Real Time Applications, pp. 141–146. IEEE (2006)
8. Qaisar, S., Bilal, R.M., Iqbal, W., Naureen, M., Lee, S.: Compressive sensing: from theory to applications, a survey. *J. Commun. Netw.* **15**(5), 443–456 (2013)
9. Sharma, S.K., Patwary, M., Abdel-Maguid, M.: Spectral efficient compressive transmission framework for wireless communication systems. *IET Signal Process.* **7**(7), 558–564 (2013)
10. Elzanati, A.M., Abdelkader, M.F., Seddik, K.G., Ghuniem, A.M.: Collaborative compressive spectrum sensing using kronecker sparsifying basis. In: 2013 IEEE Wireless Communications and Networking Conference (WCNC), pp. 2902–2907. IEEE (2013)
11. Qin, Z., Gao, Y., Plumbley, M.D., Parini, C.G.: Wideband spectrum sensing on real-time signals at sub-Nyquist sampling rates in single and cooperative multiple nodes. *IEEE Trans. Signal Process.* **64**(12), 3106–3117 (2015)
12. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. *SIAM Rev.* **43**(1), 129–159 (2001)
13. Candes, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**(8), 1207–1223 (2006)
14. Marini, F., Walczak, B.: Particle swarm optimization (PSO). A tutorial. *Chemom. Intell. Lab. Syst.* **149**, 153–165 (2015)
15. Rashedi, E., Nezamabadi-Pour, H., Saryazdi, S.: GSA: a gravitational search algorithm. *Inf. Sci.* **179**(13), 2232–2248 (2009)

16. Mirjalili, S., Mirjalili, S.M., Lewis, A.: Grey wolf optimizer. *Adv. Eng. Softw.* **69**, 46–61 (2014)
17. Faris, H., Aljarah, I., Al-Betar, M.A., Mirjalili, S.: Grey wolf optimizer: a review of recent variants and applications. *Neural Comput. Appl.* **30**(2), 413–435 (2017). <https://doi.org/10.1007/s00521-017-3272-5>
18. Melin, P., Castillo, O., Kacprzyk, J. (eds.): *Nature-Inspired Design of Hybrid Intelligent Systems*. SCI, vol. 667. Springer, Cham (2017). <https://doi.org/10.1007/978-3-319-47054-2>
19. Eltabie, O.M., Abdelkader, M.F., Ghuniem, A.M.: Incorporating primary occupancy patterns in compressive spectrum sensing. *IEEE Access* **7**, 29096–29106 (2019)
20. Baraniuk, R.G., Davenport, M.A., DeVore, R.A., Wakin, M.B.: A simple proof of the restricted isometry property for random matrices. *Constr. Approx.* **28**(3), 253–263 (2007)
21. Candes, E.J., Tao, T.: Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inf. Theory* **52**(12), 5406–5425 (2006)
22. Boyd, S., Boyd, S.P., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
23. Al-Tashi, Q., Md Rais, H., Abdulkadir, S.J., Mirjalili, S., Alhussian, H.: A review of grey wolf optimizer-based feature selection methods for classification. In: Mirjalili, S., Faris, H., Aljarah, I. (eds.) *Evolutionary Machine Learning Techniques*. AIS, pp. 273–286. Springer, Singapore (2020). https://doi.org/10.1007/978-981-32-9990-0_13
24. Bozorg-Haddad, O. (ed.): *Advanced Optimization by Nature-Inspired Algorithms*. SCI, vol. 720. Springer, Singapore (2018). <https://doi.org/10.1007/978-981-10-5221-7>



An Efficient Meta-Heuristic Methods for Travelling Salesman Problem

Mohamed Abid, Said El Kafhali[✉] , Abdellah Amzil, and Mohamed Hanini

Faculty of Sciences and Techniques, Computer, Networks, Modeling, and Mobility
Laboratory (IR2M), Hassan First University of Settat, 26000 Settat, Morocco
{mo.abid,said.elkafhali,a.amzil,mohamed.hanini}@uhp.ac.ma

Abstract. Over the past several decades, researchers have increasingly attempted to create an autonomous problem solver that might address problems in computer science, mathematics, economics, and engineering. When faced with a problem, people often go to nature for inspiration. Intelligent multi-agent systems have been inspired by the collective behavior of social insects like ants and bees, as well as other animals, such as bird flocking and fish schooling. Solutions to NP-hard issues, including the Traveling Salesman Problem (TSP), may be found through the application of algorithms such as Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), Genetic Algorithm (GA), Simulated Annealing (SA), and Tabu Search (TS). The TSP is a classic example of an NP-hard problem that has received a great deal of attention from researchers. Numerous mathematical models, software implementations, and methodological proposals have been made for TSP. TSP has been the subject of several exact and metaheuristic methods. In this research, we applied six effective metaheuristic algorithms to solve seven benchmark TSPs, including Bays29, att48, eil51, berlin52, st70, pr76 and kroa100. Using the identical settings for each simulation, we assessed the empirical data that existed in a certain arrangement. We illustrate the performance of optimal and metaheuristic solutions for TSP. ABC is shown to be near-optimal with only 1.5% degradation.

Keywords: Traveling salesman problem · Metaheuristic · ACO · PSO · ABC · GA · SA · TS

1 Introduction

The optimization problem is broadly dispersed across fields and continues to be a serious difficulty in artificial computation [1]. In the real world, challenges are classified into two types: continuous and combinatorial optimization problems. The TSP is a classic combinatorial optimization problem in which one seeks to determine the optimal path between a set of cities, whereby each city may visit

the others precisely once before returning to its starting point. The problem has been shown to be an NP-hard problem [2], which is easy to define but difficult to solve. Furthermore, many practical problems include this type of challenge, such as scheduling problems [3–5], vehicle routing problems (VRP) [6], physical mapping problems, design of integrated circuits problems, constructing phylogenetic trees problems, transportation and delivery, task allocation, tourist implementation, warehouse operations, etc. These are all examples of engineering applications that include this type of challenge. This may be generalized as TSP [7].

For TSP, many approaches have been developed, which may be described as follows: (1) exact algorithms, which include dynamic programming, branch and bound, integer linear programming, and cutting plane; (2) local search algorithms, which include the 2-Opt and 3-Opt algorithm (3) ant colony optimization algorithms (ACO) [8], artificial bee colony algorithms (ABC) [9], genetic algorithms (GA) [10], simulated annealing (SA) [11], Tabu search (TS) [12], and particle swarm optimization (PSO) [13].

The work is motivated by the following considerations: First, intensive research on the TSP has led to the development of many very powerful solvers. Second, we compare six optimization methods for the traveling salesman problem and almost all optimization problems: metaheuristics by using ACO, PSO, ABC, GA, SA, and TS.

All of these methods will be tested in Bays29, att48, eil51, berlin52, st70, pr76 and kroa100 using the MATLAB software and TSPLIB benchmark [14] data based on algorithm performance (optimality to achieve the shortest distance); the results of each method will then be statistically compared with the Matlab solver.

Specifically, this paper will follow these guidelines: In Sect. 2, we examine previous works that have addressed the ACO, PSO, ABC, TS, SA, GA, and TSP. The findings and explanations of the comparison of all existing TSP-solving algorithms are presented in Sect. 3. We conclude the paper in Sect. 4.

2 Overview and Related Work

2.1 Ant Colony Optimization

The ant colony approach arose from a simple observation: social creatures, particularly ants, solve complicated problems spontaneously. This behavior is feasible because ants communicate with one another indirectly by depositing chemical molecules on the ground known as pheromones. Stigmergy is a term used to describe this form of indirect communication. In fact, if a barrier is placed in the route of the ants, the latter will prefer to follow the shortest path between the nest and the obstruction after a period of investigation. The greater the amount of pheromone in a specific location, the more likely an ant will be drawn to that area. The ants that got to the nest the fastest through the food source used the shortest branch of the path. It follows that the number of pheromones on this path is greater than on the longest path. Therefore, the shortest path has

a higher probability of being taken by the ants than the other paths, and it will be taken by all the ants. Principle The first algorithm based on this analogy was proposed in 1996 [15]. The initial goal of this algorithm was to solve the traveling salesman problem. If we consider an N-city traveling salesman problem, each ant k traverses the graph and builds a path of length $n = |N|$. For each ant, the path from a city i to a city j depends on :

- When ant k is now on city i , the list of previously visited cities that specifies the potential motions at each step at that time is J_i^k ;
- Visibility is defined as the reciprocal of the $\eta_{ij} = \frac{1}{d_{ij}}$ distance between cities. This information is used to send the ants to neighboring cities, avoiding long journeys. The quantity of pheromone deposited on the edge linking two cities τ_{ij} , also known as the intensity of the trail. This quantity defines the attractiveness of a track, and it is modified after the passage of an ant. The moving rule is as follows:

$$p_{ij}^k(t) = \begin{cases} \frac{(\tau_{ij}(t))^\alpha (\eta_{ij})^\beta}{\sum_{l \in J_i^k} (\tau_{il}(t))^\alpha (\eta_{il})^\beta} & \text{si } j \in J_i^k \\ 0 & \text{si } j \notin J_i^k \end{cases} \quad (1)$$

in which α and β are two factors that influence the relevance of intensity and visibility. After a complete lap, each ant deposits a quantity of pheromone $\Delta\tau_{ij}^k(t)$ on its entire path. This quantity depends on the quality of the solution found and is defined by

$$\Delta\tau_{ij}^k(t) = \begin{cases} \frac{Q}{L^k(t)} & \text{if } (i, j) \in T^k(t) \\ 0 & \text{if } (i, j) \notin T^k(t) \end{cases} \quad (2)$$

where $T^k(t)$ is the path taken by the ant k at iteration t , $L^k(t)$ is the length of $T^k(t)$ and Q is a fixed parameter.

Finally, it is necessary to introduce a pheromone evaporation process. Indeed, to avoid being trapped in suboptimal solutions, it is necessary that an ant “forgets” the bad solutions. The update rule is, therefore,

$$\tau_{ij}(t+1) = (1 - \rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}(t) \quad (3)$$

where $\Delta\tau_{ij}(t) = \sum_{k=1}^m \Delta\tau_{ij}^k(t)$ and m is the number of ants. Since the first method based on the ant colony analogy was developed, this method has been used to solve both discrete and continuous optimization problems.

2.2 Artificial Bee Colony (ABC)

Many academics have wondered about the honeybee colony’s foraging behavior. We created a model using the reaction-diffusion equations [16] that predicts

the development of hive intelligence. Two dominant behavioral patterns, recruitment to a nectar supply and desertion of a source, are key to the three basic components of this model: food sources, hired foragers, and jobless foragers [16].

Foraging bees use the ABC method to choose a collection of prospective food sources at random. communicate this knowledge to observers in the hopes of enlisting them and then return to their original selection. Equation 4 [16] may be used to determine the chance p_i , that a given food source would be exploited by the hired observers, who are transformed into bees based on the probability value associated with the food source.

$$p_i = \frac{f_i}{\sum_{n=1}^{FS} f_n} \quad (4)$$

f_i stands for the fitness value of the i th response, and the quantity of food providers, denoted by FS , is equal to the number of people participating in the competition or watching it. The Eq. 5 [16] is used by ABC to produce a potential food location from a previously stored one in memory.

$$v_{ij} = x_{ij} + \phi_{ij} (x_{ij} - x_{kj}) \quad (5)$$

while, $k \in \{1, 2, \dots, FS\}$ and $j \in \{1, 2, \dots, D\}$ randomly selected; ϕ_{ij} is a number chosen at random from the interval $[-1, 1]$, and x_{ij} is the food location, also known as the solution. The number of decision variables is denoted by D in this context. At each iteration, the optimal solution is chosen by comparing all of the potential solutions, or v_{ij} , to the current solution, or x_{ij} . Even when a location cannot be upgraded after a specific number of cycles, the nectar is abandoned in favor of a food source and the scouts replace it with a fresh food source. The phrase “limit for abandoning” describes this fixed number of iterations. Equation 6 [16] states that the scout will find a new food source for an abandoned source x_i .

$$x_i^j \leftarrow x_{\min}^j + \text{rand}[0, 1] (x_{\max}^j - x_{\min}^j) \quad (6)$$

where $j \in \{1, 2, \dots, D\}$. The same strategy of giving early solutions (or a reference route) may be used to solve TSP. The artificial bees learn the optimal strategy and remember it. The best tour solutions are improved with each iteration until the termination requirement is met.

2.3 Genetic Algorithm

As early as the 1960s, Holland, a researcher at the University of Michigan, began developing genetic algorithms. His first major accomplishment came in 1975, when he created adjustments in natural and artificial systems. With his study, he had the following goals in mind: 1) better understanding the natural process of adaptation, and 2) creating artificial systems with traits that mimic those of natural systems.

The TSP serves as a conceptual model for a broad variety of other issues. Since the TSP is arguably the most important issue in combinatorial optimization, it has become a popular testbed for novel combinatorial optimization techniques, including genetic algorithms (GAs).

Tsujimura et al. [17] used GA to solve TSPs, and they discovered that quite frequently they were able to find a local optimal result instead of the best approximation. So, they used entropy-based GA to find a good approximation to the TSP solution.

2.4 Simulated Annealing Algorithm

Some consider the SA algorithm to be the first of all metaheuristics; it is certainly one of the few algorithms to use explicit tactics for avoiding local minima. SA was initially introduced for combinatorial optimization issues, although its roots may be traced back to statistical mechanics. To get out of a rut, the key is to accept steps that lead to solutions that are worse than the ones currently in use. The search probability for such a shift decreases as the parameter temperature increases. The SA method begins with a candidate solution, which is created from the area around the first solution (either arbitrarily or according to a predetermined rule). The decision to accept or reject is made using the Metropolis acceptance criteria [18].

2.5 Tabu Search (TS)

Fred W. Glover devised the Tabu search in 1986, and by 1989, it had been formally codified. It is a metaheuristic search strategy for addressing mathematical optimization problems by way of local search [19]. TS is a local search approach that restricts the search space to only include neighbors that are within a reachable region. On Tonga Island, natives have a term for things that they are too scared to touch: tabu. During a Tabu search, the state information you're looking for is kept in a data structure called a Tabu list. They assist in avoiding becoming stranded in an ideal location. If all neighbors are Tabu, a change that reduces the relevance of the goal function to which a standard is applied, the deepest decent technique would be jammed in space. The ambition criteria is a step forward: If there is a step in the Tabu list that improves all previous solutions, the Tabu limitations go undetected. [19] As a single technique for updating the Tabu list, every state visited during the preceding k phase should be prohibited. Another option is to enable local changes to not notify the same portions of the state route or to change the search cost function. The Tabu search may be regarded as a simulated annealing technique. This balances the successor set reduction by a list of tabs with a simulated achievement selection procedure. Instead of the Boltzmann state, the successors should consider [19].

2.6 Particle Swarm Optimization (PSO)

PSO has seen extensive usage for the resolution of issues involving continuous amounts [20]. It mimics the cooperative nature of animal societies like bird flocks

and fish schools. Particles are deployed in the search region to calculate the optimal solution for a specific task. Particles navigate the search space according to their own will [20]. All the particles do this operation over and over again with each cycle. The goal of particle swarm optimization (PSO) is to provide an optimization method in which these particles converge to the best possible solution.

Each particle, denoted by j , in PSO is represented by a K vector, where K is the number of dimensions in the search domain [20]. The present location of the particle j is one of these factors. $Y_j = (y_{j1}, y_{j2}, \dots, y_{jK})$, the previous best position $p_{\text{best}} = (p_{j1}, p_{j2}, \dots, p_{jK})$, and the velocity $V_i = (v_{j1}, v_{j2}, \dots, v_{jK})$. The finest global location $g_{\text{best}} = (g_1, g_2, \dots, g_K)$ is known for every n particle. Particle swarms may change their orientation by manipulating their coordinates. Equations 7 and 8 [20] describe how the velocity and location coordinates of each particle are modified after each repetition.

$$V_{jk} \leftarrow \omega \times V_{jk} + \eta_1 \times \text{rand}() \times (p_{\text{best}_{jk}} - Y_{jk}) + \eta_2 \times \text{rand}() \times (g_{\text{best}_k} - Y_{jk}) \quad (7)$$

$$Y_{jk} \leftarrow Y_{jk} + V_{jk} \quad (8)$$

with, ω denotes the weight of the inertial ($0 < \omega < 1$); η_1 as well as η_2 are constant accelerations, and rand produces a random number in the interval $[0, 1]$. The velocities of the particles are restricted to the range $[V_{\min}, V_{\max}]$.

Each particle's location Y_j and velocity V_j are calculated as part of PSO's solution to TSP. Here, j represents the particle's position and V_j its velocity. After each iteration of the j th particle visiting the cities, the results are analyzed to see how good they are and the best (p_{best}) values for each particle and the best (g_{best}) values for the entire swarm are updated accordingly.

2.7 Traveling Salesman Problem

The traveling salesman problem (TSP) is a popular and well-studied combinatorial problem in operations research. Scientists have been paying serious attention to this topic for many decades. Several exact, heuristic, and metaheuristic TSP methods have been proposed. However, there is no precise method to obtain optimal TSP solutions in polynomial time. As a consequence, finding close solutions to that problem within reasonable time constraints is possible. The fundamental formula for the TSP is [7]:

$$\text{Min. } \sum_{i=1}^n \sum_{j \neq i, j=1}^n c_{ij} x_{ij} \quad (\text{P})$$

$$\text{s.t } \sum_{i=1, i \neq j}^n x_{ij} = 1, \quad j = 1, \dots, n \quad (\text{P.a})$$

$$\sum_{j=1, j \neq i}^n x_{ij} = 1, \quad i = 1, \dots, n \quad (\text{P.b})$$

$$u_i - u_j + n x_{ij} \leq n - 1, \quad 2 \leq i \neq j \leq n, \quad (\text{P.c})$$

$$x_{ij} \in \{0, 1\} \quad (\text{P.d})$$

In the preceding formula,

1. n : the number of cities
2. c_{ij} : Assignment cost (distance) between cities i and j .
3. x_{ij} : Decision variable that equals 1 whenever there is a relationship between cities i and j .
4. u_i : are arbitrary real numbers.

The restriction (P.c) assures that each cycle is a full tour rather than a subset of whole cities. The number of viable solutions for n cities for symmetrical TSP is $(n - 1)!/2$. As a result, for situations with a medium or a high number of cities, the number of alternative solutions might be quite vast. There are various TSP instances. TSP fundamental situations include symmetric, asymmetric, and time-window scenarios. TSPLIB [14] provides versions and test data sets for the traveling salesman issue. In recent years, a wide variety of papers have been written on these issues. This section demonstrates how the ACO method may be utilized to tackle permutation-based problems, notably in the TSP. The TSP is described as visiting “ n ” cities, beginning and finishing in the same location, visiting each city just once, and completing the trip at the lowest possible cost.

3 Experiments and Results

Table 1 shows the simulation settings utilized for experimentation. In Table 2, the corresponding errors have been calculated for 7 problems using 6 methods and the Matlab solver. In all cases, the average results given by ABC are less than the corresponding average results given by ACO, PSO, GA, SA, and TS.

In Fig. 1 the performance of the optimal solution (obtained using the Matlab solver) is shown for proposed algorithms such as ACO, PSO, ABC, GA, SA and TS, in terms of the number of cities versus the distance route. The distance of the route increases for all approaches. Moreover, the Matlab solver achieves the shortest traveled distance, whereas ABC follows closely with only 0% to 1.5% degradation. The worst performance is shown by the ACO, PSO, GA, SA, and TS methods, which degrade by more than 5.6%. ABC achieves near-optimal performance.

As shown in Fig. 2, ACO converges more quickly, followed by TS and ABC. If we look at the outcomes, we can see that ABC produced superior results in previous iterations.

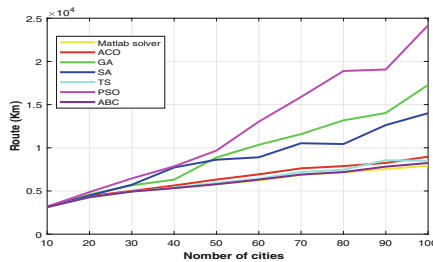


Fig. 1. Distance traveled by the TSP vs. number of cities (one run for kroA100)

Table 1. Parameter settings for TSP simulations based on a variety of methods

Parameter	ACO	PSO	ABC	GA	SA	TS
Maximum iterations	500	500	500	500	500	500
Number of agents	100 ants	100 particles	100 bees	100 chromosomes	100 Epoch length	100 Tabu list
Initial pheromone on all edges (τ_0)	1	—	—	—	—	—
Pheromone exponential weight (α)	1	—	—	—	—	—
Heuristic exponential weight (β)	5	—	—	—	—	—
Evaporation rate (ρ)	0.1	—	—	—	—	—
Pheromone quantity constant (Q)	1	—	—	—	—	—
Inertia weight	—	0.9	—	—	—	—
Acceleration constant(η_1)	—	2	—	—	—	—
Acceleration constant(η_2)	—	2	—	—	—	—
Number of onlookers	—	—	100	—	—	—
Number of scouts	—	—	25	—	—	—
Acceleration coefficient(ϕ)	—	—	[−1,1]	—	—	—
Selection strategy	—	—	—	Elitism selection	—	—
Mutation operator	—	—	—	Swap	—	—
Initial temperature	—	—	—	—	1000	—
final temperature	—	—	—	—	0.001	—
Cooling rate	—	—	—	—	0.9	—

— : Not Applicable

Table 2. Against a variety of test problems and methods

Alg.	Problem	bays29	att48	eil51	berlin52	st70	pr76	kroa100
S.Matlab	Solution	9074.148	33523.7085	428.8718	7544.3659	677.1096	108159.4383	21285.4432
GA	Average	10779.3534	47756.2534	554.9714	10198.7739	1129.9431	180865.5405	48626.7635
	Best	9921.836	42972.3449	506.6558	9008.7301	1018.7241	160102.6505	43044.485
	Error(%)	18.7919	42.4552	29.4026	35.184	66.8774	67.2212	128.4508
SA	Average	10441.2917	43919.6142	535.7185	10039.2736	1007.6022	173591.4393	40206.3466
	Best	9585.4985	37988.4829	513.6858	8808.6605	941.0561	161138.5661	34903.7923
	Error(%)	15.0664	31.0106	24.9134	33.0698	48.8093	60.4959	88.8913
TS	Average	9194.8723	34421.8106	444.3766	7990.8039	707.447	114180.5863	22487.49
	Best	9074.148	33701.2618	435.3144	7713.918	692.4079	110040.6728	21699.5009
	Error(%)	1.3304	2.679	3.6153	5.9175	4.4804	5.5669	5.6473
PSO	Average	11016.2983	47859.6326	604.0277	10406.1661	1169.3672	190448.5154	51133.5598
	Best	9733.4627	44073.246	524.6682	9249.0359	963.1169	181376.0266	46655.6462
	Error(%)	21.4031	42.7635	40.8411	37.9329	72.6998	76.0813	140.2278
ACO	Average	9359.3668	35987.0168	458.2239	7842.9105	736.7281	121287.9035	23462.4699
	Best	9316.8266	35714.3225	453.0554	7755.422	717.5313	118726.1635	23085.0632
	Error(%)	3.1432	7.348	6.844	3.9572	8.8048	12.1381	10.2278
ABC	Average	9074.148	33689.0147	432.1085	7576.7784	684.6377	109498.4911	21612.2132
	Best	9074.148	33600.5615	428.8718	7544.3659	677.1096	108159.4383	21356.5875
	Error(%)	0	0.4931	0.75472	0.42963	1.1118	1.238	1.5352

* We performed every simulation 10 times to ensure the experimentation's robustness.

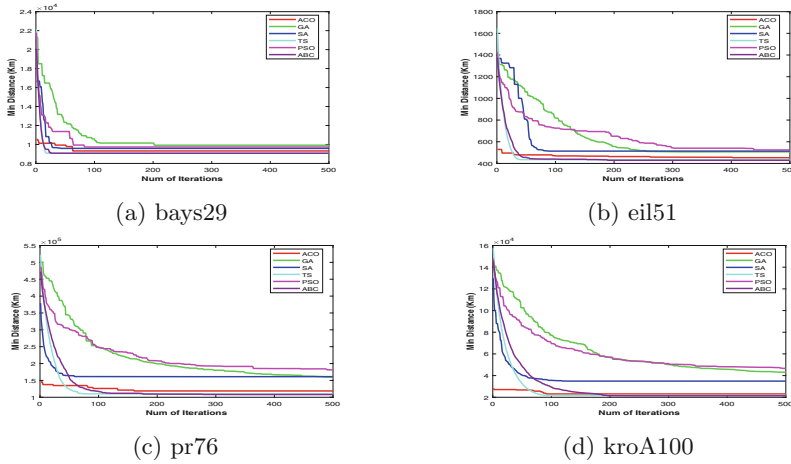


Fig. 2. Convergence curves for ACO, PSO, ABC, GA, SA and TS (best of 10 run for bays29, eil51, pr76 and kroA100)

4 Conclusions

This work conducted a study that compared the most commonly used optimization algorithm approaches (ACO, PSO, ABC, GA, SA, and TS) applied to the TSP problem. Our objective was to compare how these algorithms performed in terms of shortest distance under identical platform circumstances. Since the ABC algorithm performed best in this study's evaluations of optimization approaches applied to the TSP problem, we may deduce that it is a good option for tackling route optimization problems. Future research might benefit from combining two of the competing methods (ACO & ABC), since ABC provides the shortest distance while ACO speeds up the process. So, they balance each other out and make up for their weaknesses in other optimization problems related to the TSP.

References

1. Del Ser, J., et al.: Bio-inspired computation: Where we stand and what's next. *Swarm Evolut. Comput.* **48**, 220–250 (2019)
2. Huerta, I.I., Neira, D.A., Ortega, D.A., Varas, V., Godoy, J., Asín-Achá, R.: Improving the state-of-the-art in the traveling salesman problem: an anytime automatic algorithm selection. *Expert Syst. Appl.* **187**, 115948 (2022)
3. El Kafhali, S., El Mir, I., Salah, K., Hanini, M.: Dynamic scalability model for containerized cloud services. *Arab. J. Sci. Eng.* **45**(12), 10693–10708 (2020)
4. Hanini, M., El Kafhali, S.: Cloud computing performance evaluation under dynamic resource utilization and traffic control. In: *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications*, pp. 1–6 (2017)
5. El Kafhali, S., Hanini, M.: stochastic modeling and analysis of feedback control on the QoS VoIP Traffic in a single cell IEEE 802.16 e networks. *IAENG Int. J. Comput. Sci.* **44**(1), 19–28 (2017)

6. Roy, A., Manna, A., Kim, J., Moon, I.: IoT-based smart bin allocation and vehicle routing in solid waste management: a case study in South Korea. *Comput. Indus. Eng.* **171**, 108457 (2022)
7. Miller, C.E., Tucker, A.W., Zemlin, R.A.: Integer programming formulation of traveling salesman problems. *J. ACM (JACM)*, **7**(4), 326–329 (1960)
8. Skinderowicz, R.: Improving ant colony optimization efficiency for solving large TSP instances. *Appl. Soft Comput.* **120**, 108653 (2022)
9. Jiang, H.: Artificial bee colony algorithm for traveling salesman problem. In: 2015 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering, pp. 468–472. Atlantis Press (2015)
10. Zhang, J.: An improved genetic algorithm with 2-Opt local search for the traveling salesman problem. In: Sugumaran, V., Xu, Z., Zhou, H. (eds.) *MMIA 2021. AISC*, vol. 1385, pp. 404–409. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-74814-2_57
11. Demiral, M.F., Işık, A.H.: Simulated annealing algorithm for a medium-sized TSP data. In: Hemanth, D.J., Kose, U. (eds.) *ICAAME 2019. LNDECT*, vol. 43, pp. 457–465. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-36178-5_35
12. Khan, M.U.R., Asadujjaman, M.: A tabu search approximation for finding the shortest distance using traveling salesman problem. *IOSR J. Math.* **12**(05), 80–84 (2016)
13. Jabor, F.K., Omran, G.A., Mhana, A., Gheni, H.M.: Optimization of particle swarms for travelling salesman problem. In: 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp. 1–6. IEEE (2022)
14. Reinelt, G.: TSPLIB-A traveling salesman problem library. *ORSA J. Comput.* **3**(4), 376–384 (1991)
15. Dorigo, M., Maniezzo, V., Colorni, A.: Ant system: optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man Cybern. Part B (Cybernetics)* **26**(1), 29–41 (1996)
16. Tereshko, V., Loengarov, A.: Collective decision making in honey-bee foraging dynamics. *Comput. Inf. Syst.* **9**(3), 1 (2005)
17. Tsujimura, Y., Gen, M.: Entropy-based genetic algorithm for solving TSP. In: 1998 Second International Conference. Knowledge-Based Intelligent Electronic Systems. Proceedings KES1998 (Cat. No. 98EX111), Vol. 2, pp. 285–290. IEEE (1998)
18. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**(6), 1087–1092 (1953)
19. Edelkamp, S., Schrödl, S.: Selective search. *Heuristic Search*. Morgan Kaufmann, San Francisco, pp. 633–669 (2012)
20. Poli, R., Kennedy, J., Blackwell, T.: Particle swarm optimization. *Swarm Intell.* **1**(1), 33–57 (2007)



Optimization of Task Scheduling in Cloud Computing Using the RAO-3 Algorithm

Ahmed Rabie Fayed^(✉), Nour Eldeen M. Khalifa, M. H. N. Taha, and Amira Kotb

Department of Information Technology, Faculty of Computers and Artificial Intelligence,
Cairo University, Giza, Egypt
a.r.fayed96@gmail.com

Abstract. Cloud computing provides computer resources such as hardware and software as a service to consumers over a network. The primary idea of cloud computing is to distribute massive amounts of storage, processing, and information to scientific applications. In cloud computing, user tasks are organized and executed with appropriate resources to supply services successfully. In order to complete task scheduling, a variety of task allocation approaches are employed. This work proposes an efficient task scheduling method to improve the task scheduling approach. Optimization techniques are often used to solve NP-hard situations. This work presents an effective scheduling task method for cloud computing systems based on the RAO-3 algorithm. We evaluate the performance of our approach by applying it to three instances. According to the data, the proposed technique provided the optimal solution in terms of schedule length, speedup, efficiency, and throughput.

Keywords: Heterogeneous resources · RAO-3 algorithm · Task scheduling · Cloud computing

1 Introduction

Cloud computing gives the on-demand application access to all consumers and companies from anywhere globally. Computing becomes a service offered to customers via cloud computing. Resources are abstracted and made available to users as services based on their demand. The essential pieces that cloud providers manage are data centres. Cloud service providers charge customers for network access to resources. Instead of incurring resource investment costs, businesses obtain resources in services. As a result, customers might lower their investment in resources. Service Level Agreements (SLAs) are contracts between service providers and consumers that outline the promises that service providers provide to their customers [1]. Task scheduling distributes user tasks among resources in order to maximize usage while decreasing task execution time. There are two forms of task scheduling: dynamic scheduling and static scheduling. The scheduler is aware of task and resource specifics in static scheduling. Task and resource specifics are unknown in dynamic scheduling. The scheduler creates dynamic scheduling plans to assign user tasks to the appropriate resources. The problem of dynamic scheduling is an

NP-complete problem in which grows of the computation is exponentially in proportion to the problem size [1]. We presented an efficient solution based on the RAO-3 method termed the efficient RAO-3 algorithm (ERAO-3) to effectively tackle the task scheduling problem to decrease the schedule length and maximize the speedup, efficiency, and throughput.

The paper is organized as follows: The notations are presented in Sect. 2. Related work is presented in Sect. 3. Problem description is given in Sect. 4. The RAO-3 algorithm is given in Sect. 5. Section 6 describes the ERAO-3 approach. Section 7 presented the evaluation of the ERAO-3. Section 8 concludes and offers future work.

2 Notation

G_T	Refer to the graph of tasks
Tsk_i	Refer to the task i
V_{M_i}	Refer to virtual machine i
N_{V_M}	Refer to the virtual machine's number
N_{Tsk}	Refer to the number of tasks
$COM_Time(Tsk_i, Tsk_j)$	Refer to the communication time between Tsk_i and Tsk_j
$S_Time(Tsk_i, V_{M_j})$	Refer to the start time of task i on a V_{M_j}
$F_Time(Tsk_i, V_{M_j})$	Refer to the finish time of task i on a V_{M_j}
$R_Time(V_{M_i})$	Refer to the V_M 's ready time i
DA_LST	Refer to a list of tasks arranged in topological order of DAG
$D_Arrival(Tsk_i, V_{M_j})$	Refer to the time of task's i data arrival to V_{M_j}
$C_Time_{i,j}$	Refer to the computation time of Tsk_i in the V_{M_j}

3 Related Work

In developing computing paradigms such as cloud computing systems, scheduling is one of the most critical phases for leveraging capabilities. Cloud computing is a dynamic environment that enables services to be shared by many users. Traditional scheduling techniques are unsuitable for cloud computing systems, and this new environment needs new approaches customized to its demands. The authors of this work [2] created many methods for job scheduling in cloud computing platforms. These algorithms are based on the particle swarm optimization (PSO) algorithm, a technique inspired by animal swarms' collective and social behaviour in nature in which particles explore the problem space for an optimal or near-optimal solution. The methods were created to lower makespan, flowtime, and task execution costs simultaneously.

Cloud computing is a new computer technology that delivers dispersed, scalable, and elastic computing resources to end-users over the internet. One of the most difficult duties in the cloud computing world is task scheduling. The primary goals of task scheduling are to find the right resources for scheduling a given job on time, to make better use of the resources, and to decrease the overall completion time of all input tasks to be completed. The task scheduling issue is classified as NP-hard. Because metaheuristic algorithms are efficient in NP-hard optimization, The authors suggested a task scheduling method based on metaheuristics in this study [3]. The suggested scheduler is based on the nature-inspired grey wolf optimizer algorithm.

In the current environment, cloud computing has established itself as an emerging technology that allows organizations to use hardware, software, and applications without incurring any upfront costs over the internet. The difficulty for the cloud service provider is to manage the underlying computing resources, such as virtual machines, networks, storage units, and bandwidth, in such a way that no computer device is under-utilized or over-utilized in a dynamic environment. A good task scheduling strategy is always necessary for dynamic job allocation to avoid such a circumstance. The authors introduced the Genetic Algorithm-based job scheduling approach in this work [4], which would efficiently divide the load across the virtual machines to minimize the total response time (QoS).

For today's most demanding cloud computing service, several activities must be completed by the available resources to obtain the greatest performance, decrease reaction time, and maximize resource utilization. A novel work scheduling method that outperforms the relevant task allocation map is required to meet these problems. Because the Load Balanced Min-Min Algorithm picks the job with the shortest completion time and assigns it to the appropriate resource, it does not always create a superior makespan and does not always use resources effectively. This work [5] includes research of several task scheduling methods and a modification of the load-balanced Min-Min (ELBMM) algorithm for Static Meta-Task Scheduling. The improved approach results from a thorough examination of the impact of the load-balanced Min-Min algorithm for Static Meta-Task Scheduling in grid computing. The Enhanced load-balanced Min-Min algorithm (ELBMM) is based on the Min-Min strategy with job rescheduling to employ underutilized resources effectively. It chooses the work with the shortest completion time and assigns it to the most appropriate resource in order to provide a better timeline and make better use of resources.

Cloud computing is a new technology that allows people to pay as they go while still providing outstanding performance. Cloud computing is a heterogeneous system that holds many application data. When scheduling some intense data or computing an intensive application, it is well acknowledged that minimizing the transferring and processing time is vital to an application program. The authors develop a task scheduling model to reduce processing costs and propose a particle swarm optimization (PSO) method based on this study's [6] small position value criterion.

4 Problem Description

The task scheduling in cloud computing is represented as a Graph with N_Tsk tasks ($Tsk_1, Tsk_2, Tsk_3, \dots$, etc. Each task represents a task with G_T and E-directed edges, signifying a portion of the tasks' requests [7]. Each node represents an instruction that might be performed sequentially on the same virtual machine alongside other instructions; it contains one or more inputs. The task an exit or entry task is triggered to execute based on the availability of the inputs. A precedence-constrained partial request result ($Tsk_i \rightarrow Tsk_j$), i.e., Tsk_i precedes Tsk_j in the process of execution. The execution time of a task Tsk_i is denoted by (Tsk_i) weight. Let $COM_Time(Tsk_i, Tsk_j)$ be the time of communication of an edge, and it will be equal to zero if Tsk_i and Tsk_j are scheduled on the same virtual machine. Start and finish times are denoted by $S_Time(Tsk_i, V_M_j)$ and $F_Time(Tsk_i, V_M_j)$, respectively [7]. The $D_Arrival$ of Tsk_i at virtual machine V_M_j is given by:

$$D_Arrival(Tsk_i, V_M_j) = \max\{F_Time(Tsk_k, V_M_j) + COM_Time(Tsk_i, Tsk_k)\} \quad (1)$$

where $k = 1, 2, \dots$, number of parents.

The task scheduling issue in cloud computing may be characterized as finding the optimal assignment or schedule of the start times of the provided tasks on virtual machines. The scheduled length (completion time) and execution cost are reduced while keeping precedence constrained. The completion time is defined as the schedule length or finish time computed by:

$$Schedule\ Length = \max(F_Time(Tsk_i, V_M_j)) \quad (2)$$

$$S_Time(Tsk_i, V_M_j) = \max\{R_Time(V_M_j), D_Arrival(Tsk_i, V_M_j)\} \quad (3)$$

$$F_Time(Tsk_i, V_M_j) = S_Time(Tsk_i, V_M_j) + C_Time(Tsk_i, V_M_j) \quad (4)$$

$$R_Time(V_M_j) = F_Time(Tsk_i, V_M_j) \quad (5)$$

where $i = 1, 2, \dots, N_Tsk$, and $j = 1, 2, \dots, N_V_M$

$$Speedup = \min_{V_M_j} \left(\sum_{Tsk_i} \frac{C_Time_{i,j}}{schedule\ length} \right) \quad (6)$$

$$Efficiency = \frac{Speedup}{N_{V_M}} \quad (7)$$

$$Throughput = \frac{N_Tsk}{Schedule\ Length} \quad (8)$$

Algorithm 1: To find the schedule length [7]

```

Input the schedule of tasks
R_Time[V_Mj] = 0   where   j = 1, 2, .....N_V_M.
For i = 1 : N_Tsk
{
    From DA_LST take the first task Tski to be executed and remove it from DA_LST.
    For j = 1 : N_V_M
    {
        If Tski is scheduled to virtual machine V_Mj
        Compute start time by using Eq. (3)
        Compute finish time by using Eq. (4)
        Compute ready time by using Eq. (5)

    End If
    }
}
Calculate schedule length by using Eq. (2)
  
```

5 RAO-3 Algorithm

Let $Q(z)$ be the objective function that has to be minimized (or maximized). Assume that there are 'd' the number of design variables and 's' number of candidate solutions at iteration i (i.e. population size, $k = 1, 2, \dots, s$). Let the best candidate best acquire the best value of $Q(z)$ (i.e. $Q(z)_{\text{best}}$) in all candidate solutions, and the worst candidate worst obtain the worst value of $Q(z)$ (i.e. $Q(z)_{\text{worst}}$) in all candidate solutions. If $Z_{j,k,i}$ is the value of the j^{th} variable for the k^{th} candidate during the i^{th} iteration, then this value is adjusted according to the equation below [8]

$$Z'_{j,k,i} = Z_{j,k,i} + \text{rand}_{1,j,i}(Z_{j,\text{best},i} - Z_{j,\text{worst},i}) + \text{rand}_{2,j,i}(Z_{j,k,i} \text{ or } Z_{j,l,i} - (Z_{j,l,i} \text{ or } Z_{j,k,i})) \quad (9)$$

where $Z_{j,\text{best},i}$ is the variable j value for the best candidate and $Z_{j,\text{worst},i}$ is the variable j value for the worst candidate for the i^{th} iteration. $Z'_{j,k,i}$ is the updated value of $Z_{j,k,i}$, and $\text{rand}_{1,j,i}$ and $\text{rand}_{2,j,i}$ are the two random values in the range $[0, 1]$ for the j^{th} variable during the i^{th} iteration. The term $Z_{j,k,i}$ or $Z_{j,l,i}$ in Eq. (9) implies that the candidate solution k is compared to any randomly chosen candidate solution l , and information is exchanged depending on their fitness scores. If the fitness value of the k^{th} solution is greater than the fitness value of the l^{th} solution, the expression " $Z_{j,k,i}$ or $Z_{j,l,i}$ " is abbreviated to $Z_{j,k,i}$. If the fitness value of the l^{th} solution is greater than the fitness value of the k^{th} solution, the expression " $Z_{j,k,i}$ or $Z_{j,l,i}$ " becomes $Z_{j,l,i}$. Similarly, if the fitness value of the k^{th} solution is greater than the fitness value of the l^{th} solution, the expression " $Z_{j,l,i}$ or $Z_{j,k,i}$ " is changed to $Z_{j,l,i}$. If the fitness value of the l^{th} solution is greater than the fitness value of the k^{th} solution, the expression " $Z_{j,l,i}$ or $Z_{j,k,i}$ " is abbreviated to $Z_{j,k,i}$ [8].

RAO-3 Algorithm

Set the population size, number of design variables, and termination criterion to their initial values
Iteration=1

While iteration<= termination criterion

 Determine the best and worst solutions for the population

 Modify the solution by using Eq. (9)

 If the solution of $Z'_{j,k,i}$ better than $Z_{j,k,i}$

 Replace the old solution with the new solution that was acquired

 End if

 Iteration =iteration+1

End while

6 The Proposed Algorithm

Because the RAO-3 algorithm's vector representation is obviously in continuous value form, we will employ the five ways to transform these continuous values into discrete values. The first rule is the rule of the lowest position value (SPV) [9], the second is the Largest Position Value (LPV) [10], the third is the round nearest function, the fourth is the floor nearest function, and the fifth is the ceiling nearest function. Table 1 shows how the modulus function will be used in conjunction with the number of virtual machines in the SPV and LPV to raise the result by one.

Table 1. Convert continuous values to discrete values

Population	2.0	2.9	2.3	1.6	1.8	2.5	2.1
SPV rule	4	5	1	7	3	6	2
Modulus with SPV and $N_V_M = 3$	2	3	2	2	1	1	3
LPV rule	2	6	3	7	1	5	4
Modulus with LPV and $N_V_M = 3$	3	1	1	2	2	3	2
Round nearest function	2	3	2	2	2	3	2
Floor nearest function	2	2	2	1	1	2	2
Ceil nearest function	2	3	3	2	2	3	3

Algorithm 2: The function that converts a continuous value to a discrete value

```

Function convert_discrete(V)
RND = random number between [1 : 5]
If (RND == 1)
    use the SPV rule to convert a continuous value
Else if (RND == 2)
    use the LPV rule to convert a continuous value

Else if (RND == 3)
    use the round nearest function to convert a continuous value
Else if (RND == 4)
    use the floor nearest function to convert a continuous value
Else
    use ceil nearest function to convert a continuous value
End if

End function
  
```

Algorithm 3: ERAO-3

```

Input the directed acyclic graph with communication time and computation time

Set the population size, number of design variables, and termination criterion to their initial values

Convert the initial population by using Algorithm 2

Calculate the schedule length for the initial population by using Algorithm 1

Iteration=1

While iteration<= termination criterion
    Determine the best and worst solutions for the population

    Modify the solution by using Eq. (9)

    Convert the obtained solution  $Z'_{j,k,i}$  by using Algorithm 2

    Calculate the schedule length for the obtained solution  $Z'_{j,k,i}$  after converting by using
    Algorithm 1

    If the schedule length of  $Z'_{j,k,i}$  is better than the schedule length of  $Z_{j,k,i}$ 
        Replace the old solution with the new solution that was acquired

    End if

    Iteration =iteration+1

End while
  
```

7 Evaluation of the ERAO-3

We demonstrate the ERAO-3 performance by using it in three different scenarios. The first scenario has eight tasks and three heterogeneous virtual machines, and the second scenario includes eleven tasks and three heterogeneous virtual machines. The third is made up of two heterogeneous virtual machines and ten tasks. We set the initial values of the parameters number of population = 100, maximum number of iteration = 100.

7.1 Case 1

We consider a case of eleven tasks $\{Tsk_0, Tsk_1, Tsk_2, Tsk_3, Tsk_4, Tsk_5, Tsk_6, Tsk_7\}$ to be executed on three heterogeneous virtual machines $\{V_M_1, V_M_2, V_M_3\}$. The cost of performing every task on different virtual machines is shown in Table 2 [11]. Table 3 represents each task's start time and finish time on different virtual machines and the schedule obtained by ERAO-3. The results obtained by the ERAO-3 are compared with those obtained by micro genetic algorithm (MicroGA) [12], and genetic algorithm (GA) [7]. The results obtained by the ERAO-3, MicroGA, and GA are illustrated in Table 4. Figure 1, Fig. 2, Fig. 3, and Fig. 4 represent the results obtained by the ERAO-3, MicroGA, and GA in terms of schedule length, speedup, efficiency, and throughput.

Table 2. Computation time for case 1

Tsk/V_M	V_M1	V_M2	V_M3
Tsk ₀	11	13	9
Tsk ₁	10	15	11
Tsk ₂	9	12	14
Tsk ₃	11	16	10
Tsk ₄	15	11	19
Tsk ₅	12	9	5
Tsk ₆	10	14	13
Tsk ₇	11	15	10

Table 3. Schedule obtained by ERAO-3 for case 1

	V_M ₁		V_M ₂		V_M ₃	
	S_Time	F_Time	S_Time	F_Time	S_Time	F_Time
Tsk ₀	–	–	–	–	0	9
Tsk ₁	20	30	–	–	–	–
Tsk ₂	–	–	26	38	–	–
Tsk ₃	–	–	–	–	28	38
Tsk ₄	–	–	–	–	9	28
Tsk ₅	–	–	–	–	51	56
Tsk ₆	–	–	–	–	38	51
Tsk ₇	–	–	–	–	56	66

Table 4. The comparative results for case 1

Algorithm	MicroGA	GA	ERAO-3
Schedule length	68	68	66

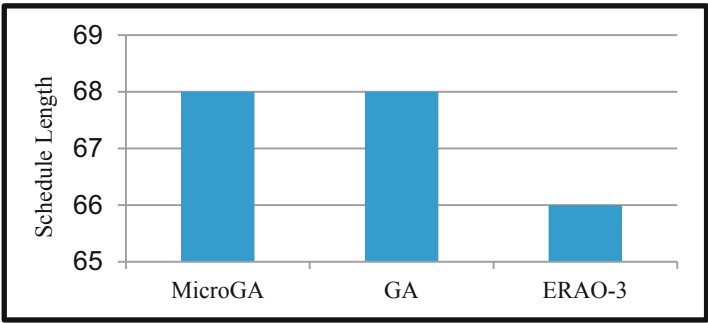


Fig. 1. Comparison of schedule length for case 1

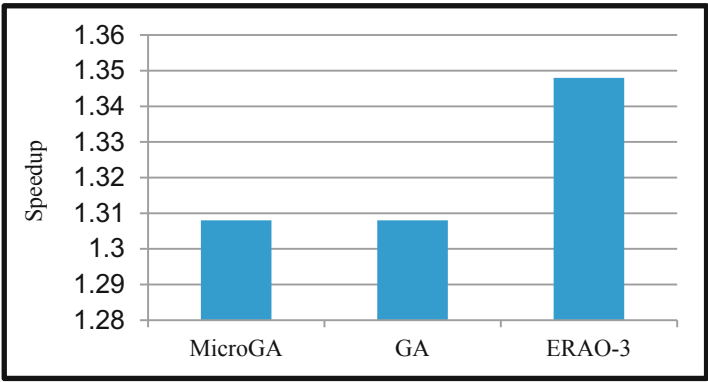


Fig. 2. Comparison of speedup for case 1

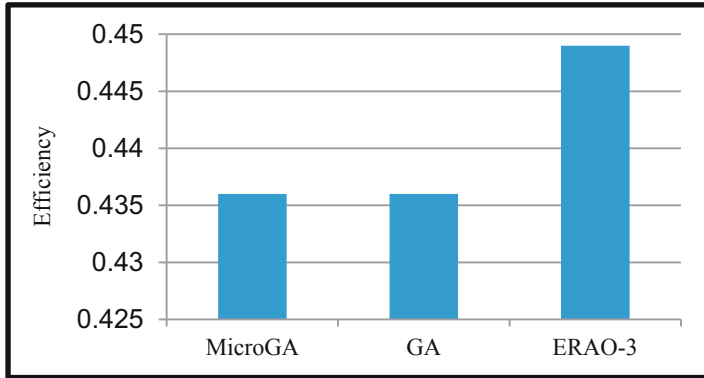


Fig. 3. Comparison of efficiency for case 1

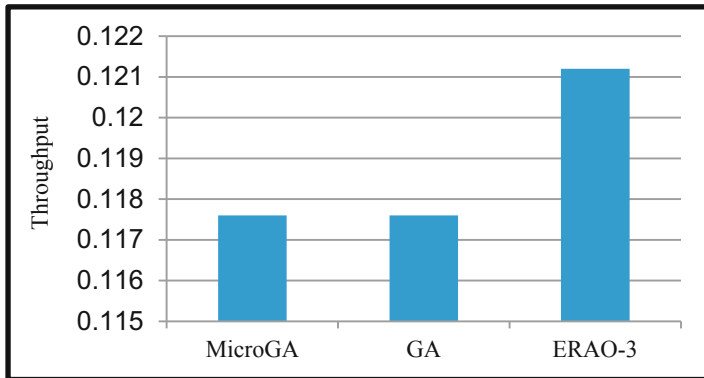


Fig. 4. Comparison of throughput for case 1

7.2 Case 2

We consider a case of eleven tasks $\{\text{Tsk}_0, \text{Tsk}_1, \text{Tsk}_2, \text{Tsk}_3, \text{Tsk}_4, \text{Tsk}_5, \text{Tsk}_6, \text{Tsk}_7, \text{Tsk}_8, \text{Tsk}_9, \text{Tsk}_{10}\}$ to be executed on three heterogeneous virtual machines $\{\text{V_M}_1, \text{V_M}_2, \text{V_M}_3\}$. The cost of performing every task on different virtual machines is shown in Table 5 [13]. Table 6 represents each task's start time and finish time on different virtual machines and the schedule obtained by ERAO-3. The results obtained by the ERAO-3 are compared with those obtained by the enhanced genetic algorithm for task scheduling (EGA-TS) [14] and genetic algorithm (GA) [7]. The results obtained by the ERAO-3, EGA-TS, and GA are illustrated in Table 7. Figure 5, Fig. 6, Fig. 7, and Fig. 8 represent the results obtained by the ERAO-3, EGA-TS, and GA in terms of schedule length, speedup, efficiency, and throughput.

Table 5. Computation time for Case 2

Tsk/V_M	V_M1	V_M2	V_M3
Tsk ₀	10	11	12
Tsk ₁	11	12	13
Tsk ₂	12	8	13
Tsk ₃	14	10	18
Tsk ₄	27	20	19
Tsk ₅	15	12	18
Tsk ₆	9	14	19
Tsk ₇	19	12	14
Tsk ₈	14	10	15
Tsk ₉	15	12	15
Tsk ₁₀	18	10	17

Table 6. Schedule obtained by ERAO-3 for case 2

	V_M ₁		V_M ₂		V_M ₃	
	S_Time	F_Time	S_Time	F_Time	S_Time	F_Time
Tsk ₀	–	–	0	11	–	–
Tsk ₁	36	47	–	–	–	–
Tsk ₂	–	–	11	19	–	–
Tsk ₃	–	–	19	29	–	–
Tsk ₄	–	–	29	49	–	–
Tsk ₅	47	62	–	–	–	–
Tsk ₆	–	–	–	–	29	48
Tsk ₇	–	–	59	71	–	–
Tsk ₈	–	–	49	59	–	–
Tsk ₉	–	–	71	83	–	–
Tsk ₁₀	–	–	83	93	–	–

Table 7. Comparative results for case 2

Algorithm	EGA-TS	GA	ERAO-3
Schedule length	104	95	93

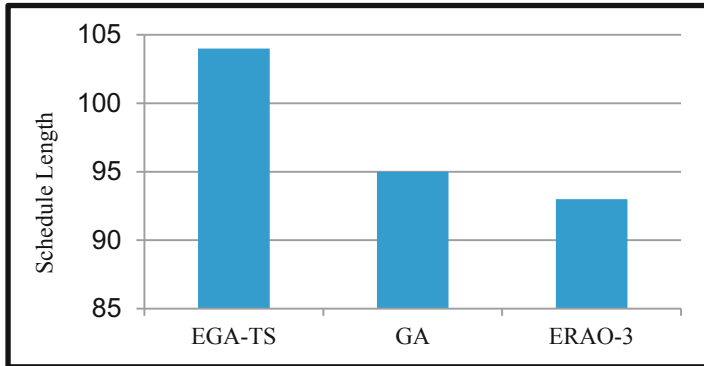


Fig. 5. Comparison of schedule length for case 2

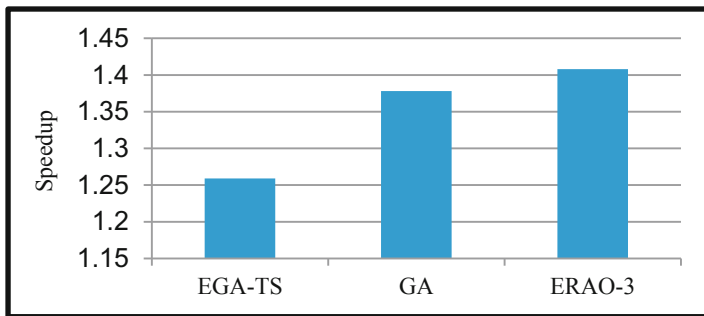


Fig. 6. Comparison of speedup for case 2

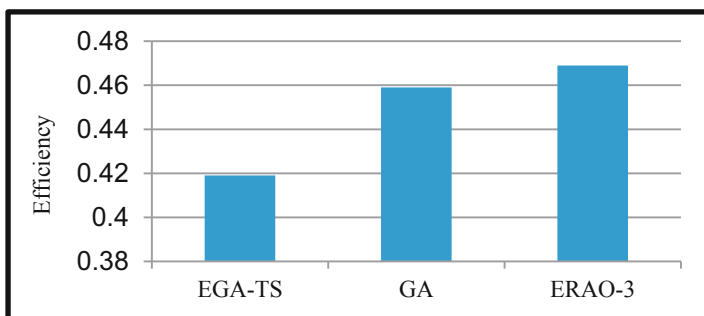


Fig. 7. Comparison of efficiency for case 2

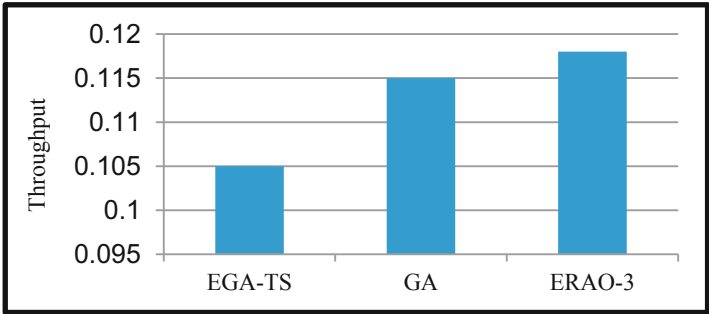


Fig. 8. Comparison of throughput for case 2

7.3 Case 3

We consider a case of eleven tasks {Tsk₁, Tsk₂, Tsk₃, Tsk₄, Tsk₅, Tsk₆, Tsk₇, Tsk₈, Tsk₉, Tsk₁₀} to be executed on two heterogeneous virtual machines {V_M₁, V_M₂, V_M₃}. The cost of performing every task on different virtual machines is shown in Table 8 [15]. Table 9 represents each task’s start time and finish time on different virtual machines and the schedule obtained by ERAO-3. The results obtained by the ERAO-3 are compared with those obtained by New Genetic Algorithm (N-GA) [13] and Proposed Particle Swarm Optimization (PPSO) [16]. The results obtained by the ERAO-3, N-GA, and PPSO are illustrated in Table 10. Figure 9, Fig. 10, Fig. 11, and Fig. 12 represent the ERAO-3, N-GA, and PPSO results in terms of schedule length, speedup, efficiency, and throughput.

Table 8. Computation cost for case 3

Tsk/V_M	V_M ₁	V_M ₂	V_M ₃
Tsk1	7	39	4
Tsk2	7	18	18
Tsk3	17	29	48
Tsk4	6	17	28
Tsk5	46	17	22
Tsk6	35	43	27
Tsk7	6	35	16
Tsk8	20	40	41
Tsk9	31	33	16
Tsk10	41	12	47

Table 9. Schedule obtained by ERAO-3 for case 3

	V_M1		V_M2		V_M3	
	S_Time	F_Time	S_Time	F_Time	S_Time	F_Time
Tsk ₁	0	7	–	–	–	–
Tsk ₂	30	37	–	–	–	–
Tsk ₃	13	30	–	–	–	–
Tsk ₄	7	13	–	–	–	–
Tsk ₅	–	–	48	65	–	–
Tsk ₆	–	–	–	–	33	60
Tsk ₇	37	43	–	–	–	–
Tsk ₈	–	–	79	119	–	–
Tsk ₉	–	–	–	–	74	90
Tsk ₁₀	–	–	123	135	–	–

Table 10. Comparative results for case 3

Algorithm	N-GA	PPSO	ERAO-3
Schedule length	175	154	135

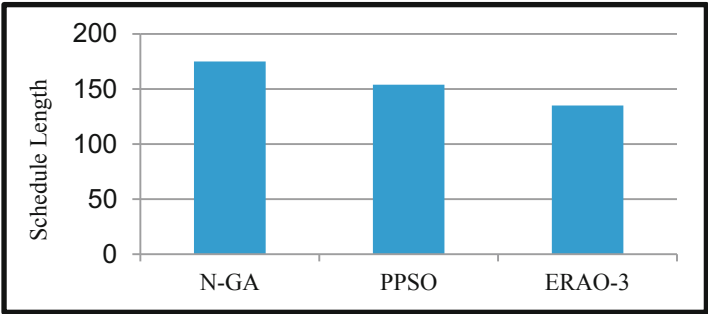


Fig. 9. Comparison of schedule length for case 3

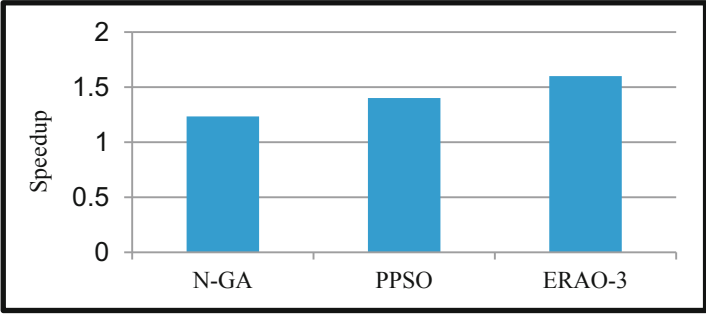


Fig. 10. Comparison of speedup for case 3

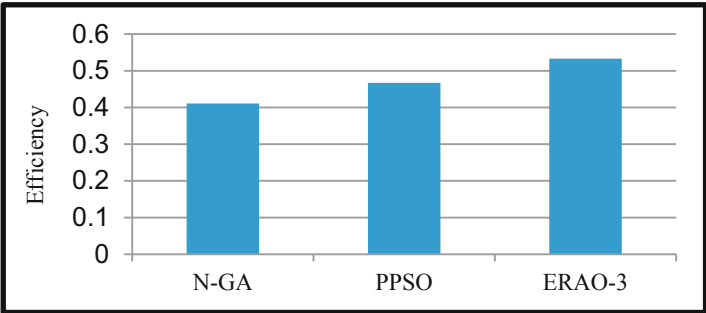


Fig. 11. Comparison of efficiency for case 3

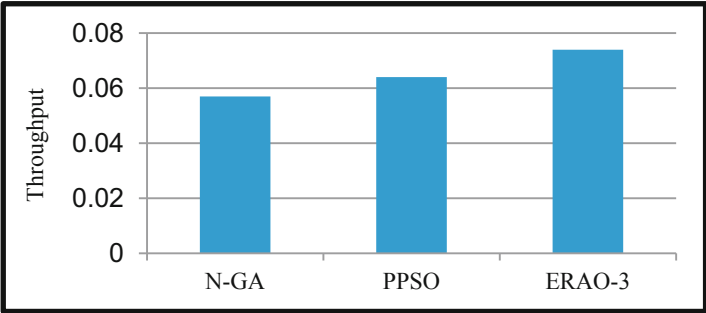


Fig. 12. Comparison of throughput for case 3

8 Conclusion and Future Work

In a cloud computing setting, the proposed efficient RAO-3 approach allocates or schedules subtasks to available virtual machines. The efficient RAO-3 method surpasses other algorithms in terms of schedule length, speedup, efficiency, and throughput, according to results collected on DAGs in various scenarios. In the future, we will develop a DAG-based method for resource load balancing.

References

1. Senthil Kumar, A.M., Venkatesan, M.: Task scheduling in a cloud computing environment using HGPSO algorithm. *Clust. Comput.* **22**(1), 2179–2185 (2018). <https://doi.org/10.1007/s10586-018-2515-2>
2. Chalack, V.A., Razavi, S.N., Gudakahriz, S.J.: Resource allocation in cloud environment using approaches based particle swarm optimization. *Int. J. Comput. Appl. Technol. Res.* **6**, 87–90 (2017). <https://doi.org/10.7753/ijcatr0602.1003>
3. Bacanin, N., Bezdan, T., Tuba, E., Strumberger, I., Tuba, M., Zivkovic, M.: Task scheduling in cloud computing environment by grey wolf optimizer. In: 27th Telecommun Forum (TELFOR 2019) (2019). <https://doi.org/10.1109/TELFOR48224.2019.8971223>
4. Agarwal, M., Srivastava, G.M.S.: A genetic algorithm inspired task scheduling in cloud computing. In: *Proceeding - IEEE Int. Conf. Comput. Commun. Autom (ICCCA 2016)*, pp. 364–367 (2017). <https://doi.org/10.1109/CCAA.2016.7813746>
5. Patel, G., Mehta, R., Bhoi, U.: Enhanced load balanced min-min algorithm for static meta task scheduling in cloud computing. *Procedia Comput. Sci.* **57**, 545–553 (2015). <https://doi.org/10.1016/j.procs.2015.07.385>
6. Guo, L., Zhao, S., Shen, S., Jiang, C.: Task scheduling optimization in cloud computing based on heuristic algorithm. *J. Netw.* **7**, 547–553 (2012). <https://doi.org/10.4304/jnw.7.3.547-553>
7. Hamed, A.Y., Alkinani, M.H.: Task scheduling optimization in cloud computing based on genetic algorithms. *Comput. Mater. Contin.* **69**, 3289–3301 (2021). <https://doi.org/10.32604/cmc.2021.018658>
8. Rao, R.V.: Rao algorithms: three metaphor-less simple algorithms for solving optimization problems. *Int. J. Ind. Eng. Comput.* **11**, 107–130 (2020). <https://doi.org/10.5267/j.ijiec.2019.6.002>
9. Dubey, I., Gupta, M.: Uniform mutation and SPV rule based optimized PSO algorithm for TSP problem. In: *Proc. 2017 4th Int. Conf. Electron Commun. Syst. (ICECS 2017)*, vol. 17, pp. 168–172 (2017). <https://doi.org/10.1109/ECS.2017.8067862>
10. Wang, L., Pan, Q.K., Tasgetiren, M.F.: A hybrid harmony search algorithm for the blocking permutation flow shop scheduling problem. *Comput. Ind. Eng.* **61**, 76–83 (2011). <https://doi.org/10.1016/j.cie.2011.02.013>
11. Lee, Y.C., Zomaya, A.Y.: A novel state transition method for metaheuristic-based scheduling in heterogeneous computing systems. *IEEE Trans. Parallel Distrib. Syst.* **19**, 1215–1223 (2008). <https://doi.org/10.1109/TPDS.2007.70815>
12. Pecero, J.E., Bouvry, P., Huacuja, H.J.F., Villanueva, J.D.T., Zuñiga, M.A.R., Santillán, C.G.G.: Task scheduling in heterogeneous computing systems using a microGA. In: *Proc. - 2013 8th Int. Conf. P2P, Parallel, Grid, Cloud Internet Comput. 3PGCIC 2013*, pp. 618–623 (2013). <https://doi.org/10.1109/3PGCIC.2013.107>
13. Keshanchi, B., Souri, A., Navimipour, N.J.: An improved genetic algorithm for task scheduling in the cloud environments using the priority queues: formal verification, simulation, and statistical testing. *J. Syst. Softw.* **124**, 1–21 (2017). <https://doi.org/10.1016/j.jss.2016.07.006>
14. Akbari, M., Rashidi, H., Alizadeh, S.H.: An enhanced genetic algorithm with new operators for task scheduling in heterogeneous computing systems. *Eng. Appl. Artif. Intell.* **61**, 35–46 (2017). <https://doi.org/10.1016/j.engappai.2017.02.013>
15. Mishra, A., Narayan Sahoo, M., Satpathy, A.: H3CSA: a makespan aware task scheduling technique for cloud environments. *Trans. Emerg. Telecommun. Technol.* **32**, 1–20 (2021). <https://doi.org/10.1002/ett.4277>
16. Biswas, T., Kuila, P., Ray, A.K.: A novel workflow scheduling with multi-criteria using particle swarm optimization for heterogeneous computing systems. *Clust. Comput.* **23**(4), 3255–3271 (2020). <https://doi.org/10.1007/s10586-020-03085-3>



Multilevel Quantum Evolutionary Butterfly Optimization Algorithm for Automatic Clustering of Hyperspectral Images

Tulika Dutta¹✉, Siddhartha Bhattacharyya^{2,3},
and Bijaya Ketan Panigrahi³

¹ Department of Computer Science and Engineering, Presidency University,
Bangalore, India

munai.tulika@gmail.com

² Rajnagar Mahavidyalaya, Rajnagar, Birbhum, India

³ Department of Electrical Engineering, Indian Institute of Technology Delhi,
New Delhi, India

Abstract. Hyperspectral images contain a huge amount of spectral and spatial information. Automatic clustering of such images is a difficult task. Two algorithms viz., Qubit Evolutionary Butterfly Optimization algorithm and Qutrit Evolutionary Butterfly Optimization algorithm, are proposed for the automatic clustering of hyperspectral images. Implementing quantum mutation operators enhances the exploration phase and prevents it from getting stuck in local optima, a drawback of their classical version. This is the first time quantum versions of the Butterfly Optimization Algorithm are proposed and implemented on such a large dataset. The two proposed algorithms are compared to the classical Butterfly Optimization Algorithm and classical Evolutionary Butterfly Optimization Algorithms, and applied to the Xuzhou HYSPEX dataset. Statistical tests like mean, standard deviation, One-Way ANOVA test, and Tukey's Post Hoc test are performed on all the algorithms to examine their efficiency. The Correlation Based Cluster Validity Index is used as the fitness function. The quality of clustering is determined using the F , F' , and Q scores. The *qutrit* version is found to produce more optimal results, followed by the *qubit* version. The *qutrit* version converges faster with better optimal results by using a smaller population compared to the other three algorithms.

Keywords: Hyperspectral image automatic clustering · Qubit Evolutionary Butterfly Optimization · Qutrit Evolutionary Butterfly Optimization · Quantum mutation operator

1 Introduction

Hyperspectral images (HSI) contain information collected over continuous spectral bands [21]. They are captured using various techniques like space-borne

sensors, air- borne sensors or ground-based sensors [21]. Due to huge spectral information, HSIs are used in different research areas like medical diagnosis [20], natural resource explorations [6], and others. The presence of huge spectral information increases redundancy in the data and requires more resources to process them [3]. Hence, dimensionality reduction is considered an important pre-processing step. The scarcity of HSI datasets is a common drawback in remote sensing, so unsupervised processes like automatic cluster detection can enhance the usage of HSI. HSI clustering divides the pixels into different meaningful groups where each group consists of pixels belonging to one specific land-cover class or crop type grown or similar mineral-containing areas [11]. In the absence of ground truth data, determining the number of sub-groups is challenging. Hence, automatic cluster detection provides better scope in HSI studies. Cluster Validity indices (CVI) I-Index [12], and Correlation Based Cluster Validity Index (CRBI) [9] are used to validate the quality of clustering [3].

As clustering is a complex optimization problem, metaheuristic algorithms are capable of finding near-optimal solutions in stipulated time [3]. A wide variety of nature-inspired metaheuristics have been researched, like Genetic Algorithm [10], Differential Evolution [14], Ant Colony Optimization [7], and others. Butterfly optimization algorithm (BOA) [1] is a recent metaheuristic algorithm mimicking butterflies' food searching and mating behaviour. However, when applied to high-dimensionality problems, BOA faces certain limitations, like getting stuck into local optima and being unable to explore the search space [17] efficiently. In [17], an enhanced BOA was proposed by introducing evolutionary operators called the Evolutionary Butterfly Optimization Algorithm (EVBOA). Even though EVBOA produces better results, the exploration of the search space is not optimal when applied to HSI datasets.

Quantum-inspired algorithms are capable of overcoming these drawbacks in less stipulated time using fewer resources [8]. Shortcomings of different metaheuristic algorithms can be eliminated to a certain extent by applying quantum parallelism. Higher-order quantum algorithms can produce better results than binary valued logic or *qubit* logic [16]. Hence, recently researchers have focused more on developing higher order quantum inspired metaheuristics.

As the "no free lunch" theorem [8] suggests that no algorithm is capable of solving all optimization problems effectively; hence researchers still have scope for developing better algorithms. This has inspired us to develop the quantum versions of the Evolutionary Butterfly Optimization algorithm [17]. The objectives of the proposed work are:

- The main aim of this work is to find the optimal number of clusters in HSI, which is a tedious task due to the high dimensionality and redundancy present in HSI data. QbEVBOA and QtEVBOA are proposed to determine the optimal number of clusters by employing a lesser number of resources and time.

- For the first time *qubit* and *qutrit* Evolutionary Butterfly Optimization Algorithm (QbEVBOA and QtEVBOA) are proposed and implemented on HSI dataset.
- The *qubit* and *qutrit* mutation operators are introduced to enhance the exploration capabilities of the EVBOA algorithm.

The organization of the paper is as follows. Section 2 contains a brief survey of the previously accomplished works. A few important concepts like the basics of quantum computing, BOA [1] and EVBOA [17] are discussed in Sect. 3. The proposed algorithms are discussed in Sect. 4. Experimental results are discussed briefly in Sect. 5. A conclusion of the proposed work is given in Sect. 6.

2 Literature Review

Hyperspectral images are captured over a wide spectral range and contain a huge amount of redundant data. Much research has been done on HSI pre-processing. Deep neural networks have been effectively applied in HSI pre-processing as they can learn linear and nonlinear features effectively [22]. The attention mechanism is effectively utilised, which helps in focusing on important features of the input [5]. This is used in [5] to develop two band selection networks viz., Spectral-based Band Selection Network using fully connected networks and Spectral-Spatial based Band Selection convolution neural network (SbBSN).

Many optimization algorithms exist in the literature, like Genetic Algorithm [10], Differential Evolution [14], and Ant Colony Optimization [7]. Three approaches are adopted to improve the effectiveness of metaheuristic algorithms. Firstly, new, improved algorithms are introduced, like BOA [1]. Secondly, existing algorithms are modified like in [3] a quantum version of Ant Colony Optimization [7] is proposed. Thirdly, hybridization of existing algorithms is done, like EVBOA [17], where mutation operators are introduced in BOA to overcome its limitations.

Quantum-inspired metaheuristics have been developed to enhance the capabilities of existing metaheuristic algorithms. Since the higher-order quantum versions are more effective, researchers have started exploiting them. In [16], a *qutrit* version of the Genetic Algorithm is proposed. The *qutrits* were initialized with equal state amplitudes using the concept of Hadamard gate for the first time. Another improved Qutrit Genetic Algorithm was proposed in [8] where different genetic operators were used for thresholding of HSI.

3 Important Concepts

Some important background concepts like principles of quantum computing, BOA and EVBOA algorithms and CBRI indices are discussed in this section.

3.1 Basic Principles of Quantum Computing

A *qubit* is the smallest unit of a quantum computer. It is denoted using Dirac's bra and ket notations [3]. *Qubits* exist in *ground* ($|0\rangle$) or *excited* ($|1\rangle$) state. The following column vector notations are used to represent a *qubit*:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (1)$$

Qubits can also exist in a linear combination of the basis states which is known as the *quantum superposition principle*. A superposition state (ψ) can be expressed as follows:

$$|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle \quad (2)$$

In Eq. (2), α_0 and α_1 represent the probabilities of finding the *qubit* in $|0\rangle$ or $|1\rangle$ state, respectively. The normalized form of a *qubit* is as follows:

$$\alpha_0^2 + \alpha_1^2 = 1 \quad (3)$$

Recently, quantum computing has caught the attention of various researchers as it overcomes the limitations of traditional bits and can exist in more than the two basis states ($|0\rangle, |1\rangle, |2\rangle, \dots, |n\rangle$ number of states). The smallest unit of a multi-valued quantum state is called a *qutrit*. The smallest of the multi-valued logic is the *qutrit* or the three-valued state. In column vector notation, a *qutrit* is represented as follows:

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, |1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \text{ and } |2\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (4)$$

The superposition of a *qutrit* state is expressed in the following manner:

$$|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle + \alpha_2|2\rangle \quad (5)$$

$$\alpha_0^2 + \alpha_1^2 + \alpha_2^2 = 1 \quad (6)$$

Equation (6), represents the normalization principle of a *qutrit* state.

Compared to classical bits, *qubits* or *qutrits* can accomplish similar tasks in lesser time, utilising fewer resources [16]. If n *qubits* are needed to express information, then $\frac{n}{\log_2(3)}$ *qutrits* are required for the same purpose. The efficiency increases by $\log_2(3) \approx 1.6$ [16].

3.2 Evolutionary Butterfly Optimization Algorithm

BOA [1] imitates the feeding and mating behaviour of butterflies. Butterflies have chemoreceptors scattered on their bodies. They sense the availability of food, flowers, or optimal partners near them using chemoreceptors. The butterflies generate a fragrance that acts as the guiding or search agent. If a butterfly is unable to detect the fragrance, it will go for exploitation by moving into a new

search space. If it finds the fragrance, it will do exploration and navigate the present space. The fragrance is mathematically defined as follows [1]:

$$Fr = CS^a \quad (7)$$

where Fr is the magnitude of fragrance of a butterfly, C is the modality of the fragrance ranging from $[0, 1]$ and S is the intensity of the emitted fragrance. In Eq. (7), a is the power exponent, which depends on the fragrance modality and ranges between $[0, 1]$. It is also responsible for controlling the fragrance absorption by stimulus intensity. The following equations are used to update the location of a butterfly for the next iteration:

$$X_i^{t+1} = X_i^t + (rand^2 \times Gb^* - X_i^t) \times Fr_i \quad (8)$$

$$X_i^{t+1} = X_i^t + (rand^2 \times X_j^t - X_k^t) \times Fr_i \quad (9)$$

In Eqs. (8) and (9), Gb value represents the best solution of the current iteration (t), Fr_i represents the fragrance magnitude of the i^{th} butterfly and $rand$ is a random number between $[0, 1]$. X^t represents the position of the butterflies, where X_i^t is the i^{th} butterfly, X_j^t is the j^{th} butterfly and X_k^t is the k^{th} butterfly. A switch probability ρ is considered for switching between exploration and exploitation for each butterfly. In every iteration, a random number is considered. If its value is less than ρ , the concerned butterfly undergoes exploration using Eq. (8), or if its value is more, then it does exploitation using Eq. (9). In the EVBOA [17], after a certain number of iterations, if the solution does not improve, a mutation operator is introduced. If the mutated generation gives a better solution, then it is retained else a random butterfly is replaced with the current butterfly.

3.3 Correlation Based Cluster Validity Index

A novel CVI was introduced in [9]. To determine the optimal cluster numbers, the following equation is used:

$$CRBI = 1 - \frac{\frac{1}{l(l-1)} \sum_{i,j=1}^l (1 + \text{Corr}(\mathbf{c}_i, \mathbf{c}_j))}{2 + \frac{1}{l} \sum_{i=1}^l \frac{1}{|Cl_i|(|Cl_i|-1)} \times \sum_{r,s=1}^{|Cl_i|} \sum_{r \neq s}^{\mathbf{g}_r, \mathbf{g}_s \in Cl_i} (1 + \text{Corr}(\mathbf{g}_r, \mathbf{g}_s))} \quad (10)$$

where Cl_i depicts the i^{th} cluster, c_i is the cluster center and $\text{Corr}(\mathbf{c}_i, \mathbf{c}_j)$ is defined by.

$$\text{Corr}(\mathbf{c}_i, \mathbf{c}_j) = \frac{\sum_{t=1}^m (c_{it} - \bar{c}_i)(c_{jt} - \bar{c}_j)}{\left[\sum_{t=1}^m (c_{it} - \bar{c}_i)^2 \sum_{t=1}^m (c_{jt} - \bar{c}_j)^2 \right]^{\frac{1}{2}}} \quad (11)$$

where,

$$\bar{c}_i = \frac{1}{m} \sum_{t=1}^m c_{it} \quad (12)$$

and

$$\text{Corr}(\mathbf{g}_r, \mathbf{g}_s) = \frac{\sum_{t=1}^m (g_{rt} - \bar{g}_r)(g_{st} - \bar{g}_s)}{\left[\sum_{t=1}^m (g_{rt} - \bar{g}_r)^2 \sum_{t=1}^m (g_{st} - \bar{g}_s)^2 \right]^{\frac{1}{2}}} \quad (13)$$

where, $\bar{g}_r = \frac{1}{m} \sum_{t=1}^m g_{rt}$ and $\bar{g}_s = \frac{1}{m} \sum_{t=1}^m g_{st}$.

Equation (13) determines the degree of correlation between two pixels. A value of 1 indicates that the two considered pixels have a high correlation, whereas a value of 0 indicates that they are independent of each other. CRBI [9] considers intra-cluster similarity and inter-cluster dissimilarity to evaluate the clustered images. The value for CRBI ranges between $[0, 1]$, where 1 indicates optimal clustering.

4 Proposed Methodology

The total work is divided into two main parts viz., the band minimization and automatic clustering of the minimized bands using QbEVBOA and QtEVBOA.

4.1 SbBSN

For pre-processing, the SbBSN [5] architecture is used. It employs the attention mechanism to learn linear and non-linear features and reconstructs the minimised HSI with some informative bands. SbBSN consists of a band attention module, a re-weighting module, and a reconstruction network module.

The attention module extracts spectral and spatial information employing multiple 2-D convolution layers. For learning the weights, the Sigmoid function is used. A global pooling layer is used to reduce the spatial size of the feature maps. A fully connected layer is then used to reweigh the spectral bands, followed by a band re-weighting module. Finally, a reconstruction network is implemented using a convolution-deconvolution network and the informative bands are obtained by averaging the band weights.

4.2 Qubit and Qutrit Evolutionary Butterfly Optimization Algorithm

The EVBOA [17] algorithm has been modified by introducing a quantum population and modified quantum mutation operator.

The initial quantum population (RP) of (no) number of butterflies is generated using the following equations for *qubit* [8] and *qutrit* [16] respectively:

$$|RP_{qb}\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \quad (14)$$

$$|RP_{qt}\rangle = \frac{1}{\sqrt{3}}|0\rangle + \frac{1}{\sqrt{3}}|1\rangle + \frac{1}{\sqrt{3}}|2\rangle \quad (15)$$

Initially, the same values are assigned as the butterflies contain no prior information about their fragrance or the location of the food sources, representing the application of Hadamard gates [16]. The probable classical values (CL) are obtained using a random value (r) [16]. If the random values are greater than α_0^2 then probable CL is assigned to 0 else to 1 for a *qubit* [8]. For *qutrit* [16], if α_0^2 is less than r then probable CL is assigned to 0, or if r is greater than α_0^2 but greater than $\alpha_0^2 + \alpha_1^2$ then CL is assigned to 1. If both cases fail, then CL is assigned to 2. Here, $\alpha_0, \alpha_1, \alpha_2$ are the probability of the quantum states being in $|0\rangle, |1\rangle$ and $|2\rangle$ states.

The main objective of the proposed algorithms is to find the optimal number of clusters. Random non-zero values are introduced in the butterfly population, which denote the cluster numbers. The corresponding quantum values for which CL is true are considered the cluster centers. The clusters are created using a *Fuzzy C Means* algorithm [2]. The optimality of the number of clusters is determined using CBRI index [9].

After finding the fitness, the fragrance of individual butterflies is calculated using Eq. (7). Then based on a random value, as mentioned in Sect. 3.2, some butterflies are chosen for exploration and some for exploitation. For QbEVBOA, if CL value is 0, then α_0 is updated using Eq. (8) for exploration or Eq. (9) for exploitation. Simultaneously, α_1 values are updated using the normalization principle (Eq. (3)). Similarly, for QtEVBOA, based on CL value, the corresponding *alpha* value gets updated [8].

The *qubit* and *qutrit* mutation operators bring diversity in the population and prevent it from getting stuck in local optima. A random value is taken between $[0, 1]$ and if it is less than α_0^2 , then α_0 and α_1 values are interchanged for a *qubit*. For *qutrit*, firstly the same procedure as *qubits* is done. If the random value is greater than α_0^2 but less than $\alpha_0^2 + \alpha_1^2$, then α_1 and α_2 values are interchanged. If both the above cases do not satisfy then α_0 and α_2 values are interchanged. This procedure is continued until the mutation probability values are reached. The optimal number of clusters is taken for the highest CBRI value obtained.

5 Results

The Xuzhou HYSPEX dataset [15, 19] is used for comparing the results of the compared methods. It is a relatively new HSI dataset captured over the Xuzhou peri-urban site. The spatial resolution is 0.73 m/pixel. 436 bands of 500×260 pixels each, are present in the dataset, classified into 9 parts.

The *CBRI* cluster validity index is used for determining the optimal number of clusters. The two proposed algorithms, QbEVBOA and QtEVBOA are compared with classical BOA [1] and EVBOA [17] to compare their effectiveness. The initial population size is taken as 20 each for BOA [1] and EVBOA [17]. For QbEVBOA, the population size chosen is 17; for QtEVBOA, it is 14. All five algorithms are run 50 times each for 100 iterations. The codes are written in MATLAB R2019a, and processed in Intel (R) Core (TM) i7 8700 in Windows 10 environment. To obtain an impartial analysis of the performance of the compared algorithms, the same environment is used for all.

The Mean, standard deviation, best convergence time, optimum fitness values, and corresponding cluster numbers for all five algorithms are recorded in Table 1. The ground-truth image and clustered images are represented in Fig. 1. To check the optimality of the clustered images, three automatic evaluation functions, F score [4], F' score [4], and Q score [4] are used. The normalized values are recorded in Table 2.

The One-Way ANOVA test [13] is applied to verify the Null hypothesis with 1% significance level. The p value obtained is $1.5441e-46$ less than 0.0001, indicating that the result is significant. So the null hypothesis, which states that all the results from all the five methods belong to the same distribution, is rejected. The corresponding box-plot for the One-Way ANOVA test [13] is presented in Fig. 2(a). A post-hoc test called the Tukey's test [18] is also performed on the obtained results to determine which fitness values differ from one another for the participating algorithms and is presented in Fig. 2 (b).

The presented results indicate that the *gutrit* version of EVBOA [17] algorithm provides more optimum results in far lesser time by utilizing a smaller population size followed by the *qubit* version. HSI pre-processing is difficult; hence, unsupervised cluster detection is very much required.

Table 1. Mean, Standard Deviation (STD), best reported time and the optimal number of clusters (NC) detected for BOA [1], EVBOA [17], QbEVBOA, and QtEVBOA using CBRI [9] for the Xuzhou HYSPEX dataset [15,19]

Sr. no.	Algorithms	Mean	STD	Time	CBRI [9]	NC
1	BOA [1]	8.0491e-05	6.9590e-05	391.298	0.000025	06
2	EVBOA [17]	3.2800e-05	8.0201e-05	279.723	0.000441	07
3	QbEVBOA	4.8762e-04	9.4727e-06	144.621	0.000507	10
4	QtEVBOA	6.1118e-04	1.0883e-05	106.868	0.000646	10

Table 2. F [4], F' [4] and Q [4] for BOA [1], EVBOA [17], QbEVBOA, and QtEVBOA using CBRI [9] for the Xuzhou HYSPEX dataset [15,19]

Sr. no.	Algorithms	F [4]	F' [4]	Q [4]
1	BOA [1]	0.0251	0.1497	0.2419
2	EVBOA [17]	0.7039	0.6808	0.0016
3	QbEVBOA	0.4799	0.8752	0.5138
4	QtEVBOA	0.4151	0.8752	0.5138

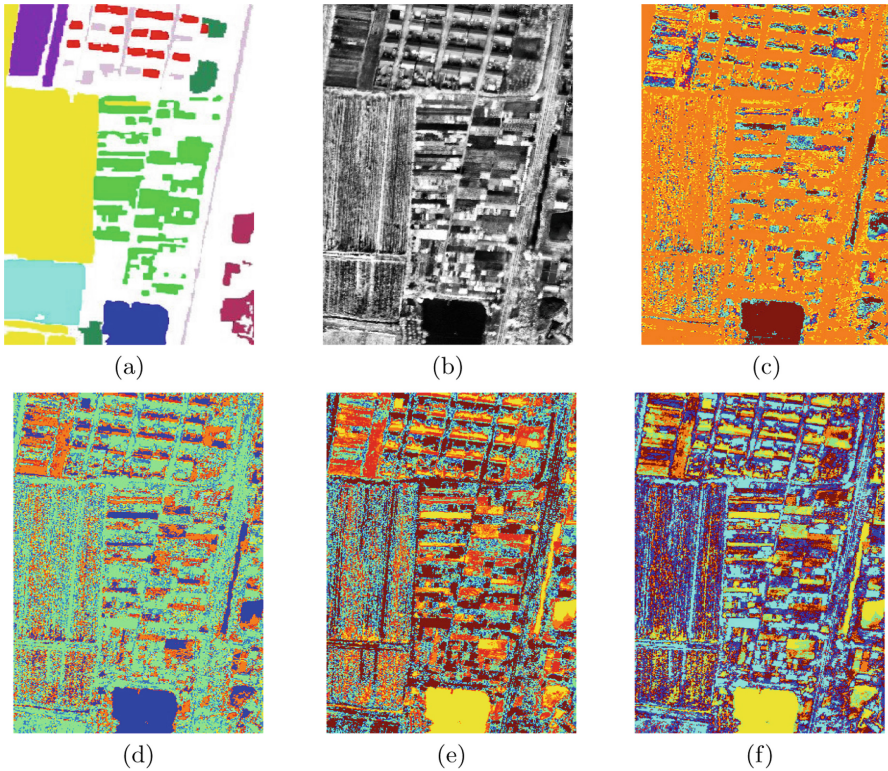


Fig. 1. (a) Ground Truth Image of the Xuzhou HYSPEX dataset [15,19], (b) Minimized Image using SbBSN [5], (c)–(f) Clustered Image using BOA [1], EVBOA [17], QbEVBOA, and QtEVBOA using CBRI [9] on the Xuzhou HYSPEX dataset [15,19]

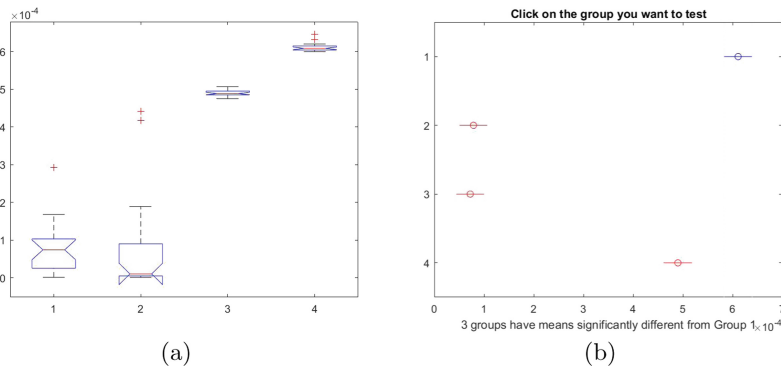


Fig. 2. (a) One-Way ANOVA Test [13] and (b) Tukey's Post Hoc Test [18] results for the Xuzhou HYSPEX dataset [15,19] using CBRI (compared with QtEVBOA) [9]

6 Conclusion

Qubit Evolutionary Butterfly Optimization algorithm and Qutrit Evolutionary Butterfly Optimization algorithm are proposed to cluster hyperspectral images automatically. Enhanced quantum mutation operators are proposed, which enhance the exploration ability of the EVBOA algorithm, thereby overcoming the drawbacks of its classical versions. The proposed algorithms are compared with the classical Butterfly Optimization Algorithm and Evolutionary Butterfly Optimization Algorithm with the help of different statistical tests. The *qutrit* version obtains more optimal results in lesser time while utilising a smaller population size. The main drawback of these algorithms is that the number of candidates chosen for exploration and exploitation is selected randomly. A dynamic system can be introduced for a smooth transition between exploration and exploitation, a future direction the authors would concentrate on.

References

1. Arora, S., Singh, S.: Butterfly optimization algorithm: a novel approach for global optimization. *Soft Comput.* **23**(3), 715–734 (2019)
2. Bezdek, J.C., Ehrlich, R., Full, W.: Fcm: the fuzzy c-means clustering algorithm. *Comput. Geosci.* **10**(2), 191–203 (1984)
3. Bhattacharyya, S., Dutta, T., Dey, S.: Multilevel quantum inspired fractional order ant colony optimization for automatic clustering of hyperspectral images. In: 2020 IEEE Congress on Evolutionary Computation (CEC), pp. 1–8 (2020)
4. Borsotti, M., Campadelli, P., Schettini, R.: Quantitative evaluation of color image segmentation results. *Pattern Recogn. Lett.* **19**(8), 741–747 (1998)
5. Cai, Y., Liu, X., Cai, Z.: Bs-nets: An end-to-end framework for band selection of hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **58**(3), 1969–1984 (2020)
6. Ding, C., et al.: Hyperspectral image classification promotion using clustering inspired active learning. *Remote Sens.* **14** (2022)
7. Dorigo, M., Maniezzo, V., Colnari, A.: Ant system: optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **26**(1), 29–41 (1996)
8. Dutta, T., Bhattacharyya, S., Mukhopadhyay, S.: Automatic clustering of hyperspectral images using qutrit exponential decomposition particle swarm optimization. In: 2021 IEEE International India Geoscience and Remote Sensing Symposium (InGARSS), pp. 289–292 (2021)
9. Ghosh, Swarup Kr, Ghosh, Anupam: Correlation based cluster validity index for recognition of leukemia mediating biomarkers. In: Mandal, Jyotsna Kumar, De, Debashis (eds.) EAIT 2021. LNNS, vol. 292, pp. 65–74. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-4435-1_8
10. Holland, J.H.: *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press, Cambridge (1992)
11. Lei, J., Li, X., Peng, B., Fang, L., Ling, N., Huang, Q.: Deep spatial-spectral subspace clustering for hyperspectral image. *IEEE Trans. Circ. Syst. Video Technol.* **31**(7), 2686–2697 (2021)

12. Maulik, U., Bandyopadhyay, S.: Performance evaluation of some clustering algorithms and validity indices. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 1650–1654 (2002)
13. Rödel, E., Fisher, R.A.: Statistical methods for research workers, 14. aufl., oliver & boyd, Edinburgh, London 1970. xiii, 362 s., 12 abb., 74 tab., 40 s. *Biometrische Zeitschrift* **13**(6), 429–430 (1971)
14. Storn, R., Price, K.: Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* **11**(4), 341–359 (1997)
15. Tan, K., Wu, F., Du, Q., Du, P., Chen, Y.: A parallel gaussian-bernoulli restricted boltzmann machine for mining area classification with hyperspectral imagery. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **12**(2), 627–636 (2019)
16. Tkachuk, V.: Quantum genetic algorithm based on qutrits and its application. *Math. Prob. Eng.* **2018**(8614073) (2018)
17. Tubishat, M., Alswaitti, M., Mirjalili, S., Al-Garadi, M.A., Alrashdan, M.T., Rana, T.A.: Dynamic butterfly optimization algorithm for feature selection. *IEEE Access* **8**, 194303–194314 (2020). <https://doi.org/10.1109/ACCESS.2020.3033757>
18. Tukey, J.W., et al.: *Exploratory data analysis*, vol. 2. Reading, MA (1977)
19. Wang, X., Tan, K., Du, Q., Chen, Y., Du, P.: Caps-triplegan: gan-assisted capsnet for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **57**(9), 7232–7245 (2019)
20. Weijtmans, P., Shan, C., Tan, T., de Koning, S., Ruers, T.J.M.: A dual stream network for tumor detection in hyperspectral images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1256–1259 (2019)
21. Zhao, J., Yan, H., Huang, L.: A joint method of spatial-spectral features and bp neural network for hyperspectral image classification. *Egypt. J. Remote Sens. Space Sci.* **26**(1), 107–115 (2023)
22. Zhong, Z., Li, J., Luo, Z., Chapman, M.: Spectral-spatial residual network for hyperspectral image classification: a 3-d deep learning framework. *IEEE Trans. Geosci. Remote Sens.* **56**(2), 847–858 (2018)



Enhancement of Security in GFDM Using Ebola-Optimized Joint Secure Compressive Sensing Encryption and Symbol Scrambling Model

Irfan Ahmad Rather¹, Gulshan Kumar^{1(✉)}, Rahul Saha¹, and Tai-hoon Kim²

¹ School of Computer Science and Engineering, Lovely Professional University, Punjab, India
gulshan3971@gmail.com

² Konkuk University, Seoul, South Korea

Abstract. The security of data transmission over the Passive Optical Network (PON) is a very challenging task. In this research work, joint secure compressive sensing encryption and symbol scrambling method is introduced in the Generalized Frequency Division Multiplexing (GFDM) systems to overcome this problem. After compression, CS data is converted into Binary Phase Shift Keying (BPSK) symbols, the SHA-256 technique is used for key generation purposes, and the key selection process is performed to improve the high security of the proposed system. Ebola meta-heuristic scheme is proposed for the optimal key selection process. The maximum Peak-To-Signal Noise Ratio (PSNR) value is the Objective Function (OF) of this proposed key selection to choose the optimal secret key from the random numbers. Finally, the symbol scrambling function is performed on the resulting BPSK symbols to improve the security of the CS data. The performance of this proposed system is compared with some of the existing methods. The achieved values of the proposed GFDM-PON for PSNR, MSE, Entropy, encryption time, and decryption time are 57.86319, 0.004114, 9.495201, 0.054, and 0.048, respectively.

Keywords: Passive optical network · Compressed sensing · Binary phase shift keying · Key selection · Peak to noise ratio

1 Introduction

The growing demand for wireless applications, as well as the rapid growth in the number of connected users, have fatigued the capacity of today's wireless communication systems [1, 2]. The Orthogonal Frequency Division Multiplexing Passive Optical Network (OFDM-PON) has proven to be a promising choice to meet the expectations of future access networks [3]. The current approach to securing such information relies heavily on encryption methods introduced at higher layers, which can be vulnerable to technologies [4]. The next fifth-generation (5G) cellular network will be necessary to offer absolute connectivity for many IoT systems [5].

A breakthrough approach of the optical OFDM method has been proposed to achieve the modulation with high spectral efficiency such as OFDM-PON [6]. Traditionally, security in communications has been established at the application layer using cryptographic encryption approaches that rely on computational resource limitations for decryption [7]. The detriment of a multi-user wireless communication system is that the MIMO has directional gain and spatial selectivity for the transmitter, but lacks for the receiver [8]. The researchers face additional hurdles if the industry moves toward 5G wireless communication [9].

Another technique designed in [10] is to consider channel estimation mistakes to construct resilient beam forming. To ensure Physical Layer Security (PLS) for multi-user beam division multiple access (BDMA) in a huge multiple-input multiple-output (MIMO) system, an OFDM-PON presents a chaos-based physically enhanced security system. It uses pseudo-random number sequence to adjust the frequency of RF subcarriers; these RF subcarrier's chaotic frequencies are difficult to decrypt by an eavesdropper due to the chaotic feature [11]. A real-valued chaotic orthogonal matrix transform-based encryption technique for increasing PLS in passive optical networks using fast OFDM [12] is also noteworthy.

Another system uses OFDM-PON device that combines the Chaotic Arnold Transform (CAT) and Chaotic Chirp Matrix (CCM) algorithm to enhance the PLS [13]. For Wavelength Division, Multiplexing Passive Optical Networks (OFDM-WDM-PON) merge Chaotic Neural Networks (CNNs). The technique enhances the key space, adjustability, and flexibility of chaotic encryption devices by integrating chaos into neural networks. It will use linear and non-linear encryption to encrypt the bit series, and subcarrier positions of the OFDM signal explained in [14, 15].

In [16], the authors propose the FrFT method to enhance the transmission performance and PLS. Because of the key generation, this method has to face processing time complexity. Further, in [17], the proposed CS based on OFDM-PON is used to improve the security of data transmission. Still, the chaotic CS method was an innovative compressive data transmission, which led to cause learning difficulties. In [18], the author implements OFDM-PON based on Brownian motion here, and the processing time minimization depends on the side length of a matrix. Such drawbacks of the existing system motivate us to develop secure compressive sensing encryption and symbol scrambling strategy for data security enhancement.

2 Proposed Methodology

Figure 1 demonstrates the framework of the proposed methodology.

a. Joint Secure Compressive Sensing Encryption

The two significant matrices in the compressed sensing skeleton are the measurement basis matrix (MBM) and the sparse basis matrix (SBM). The conventional matrices utilized to work as an SBM are discrete Fourier transform (DFT), DCT, and discrete wave transform (DWT). The chaotic matrix and the partial Hadamard are the two strategies that serve as a measurement matrix. Here, the chaotic matrix is produced by a chaotic map.

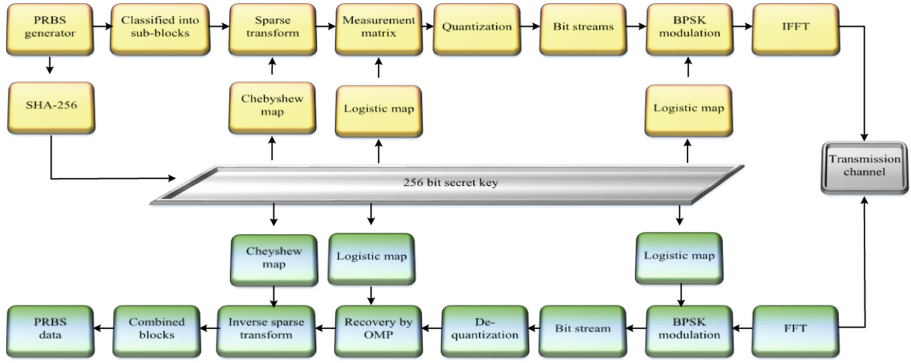


Fig. 1. Framework of the proposed methodology

Compressed Sensing. The compressed or sparse signal is obtained and reconstructed from the domain called ψ by using very small measurements besides the traditional Nyquist rate. The compressed or sparse signal is represented as $P \subset r^n$. The value P is given as follows

$$P = \psi S \quad (1)$$

where, $S \subset r^n$. The value S is assumed to be k sparse if $k(k \ll m)$ non-zero elements are present S . The size of the measurement vector is $m(m < n)$, and is said to be q is mentioned below

$$q = \Theta p + nv = \Theta \psi S + nv = Xv + nv \quad (2)$$

where the measurement matrix (MM) is Θ , $X = \Theta \psi$ is known as the sensing matrix and here nv denotes the noise vector. Θ denotes the MM, which is built by utilizing a chaotic map. Then the restricted isometry (RIP) constraint is satisfied by the matrix X .

$$1 - \xi_K \leq \frac{\|Xv\|_2}{\|v\|_2} \leq 1 + \xi_K \quad (3)$$

where, $\xi_K \in (0, 1)$. Based on expression (2), the reconstructed signal v is produced. The reconstructed method is of two types: one is convex optimization, and another is greedy techniques. The secure compressed sensing is focused in this research work so that the conventional orthogonal matching pursuit (OMP) method in terms of greedy technique is applied to learn the proposed secured strategy performance.

- 1) **Secure compressed sensing strategy:** A block compressive sensing secure strategy is used in the research work, it uses a chaotic measurement matrix and chaotic basis matrix for the binary sequences. The actual binary sequence is classified as the number of sub-blocks. Every block is converted into an SB region with the aid of chaotic DCT. Then measure the outcome signal by chaotic DCT matrix.

The actual binary sequence is classified into some sub-blocks to minimize the calculation complexity of the CS. Initially, each sub-block is converted into a chaotic DCT SB based upon the DCT row scrambling. From the Logistic map, the chaotic sequence is produced.

$$p(N + 1) = \lambda p(N)(1 - p(N)) \quad (4)$$

where the initial value is denoted as $p(0)$ and the value of $p(N) \in (0, 1)$. The expression (4) becomes a chaotic model if the value λ meets the range of $3.5699 < \lambda \leq 4$. Here, λ is the control parameter. Finally, the chaotic vector of size N_B is produced. In terms of the logistic map, the obtained chaotic vector is $P = [p(1), p(2), \dots, p(N_B)]$. The β vector's n th element is expressed as,

$$\beta(N) = \text{mod} \left(\text{floor} \left((p(N) + 100) \times 10^{10} \right), N_B \right) + 1 \quad (5)$$

The obtained new chaotic integer scrambling vector is expressed with the help of $\lambda_1 = [\lambda(1), \lambda(2), \dots, \lambda(N_B)]$. The conventional DCT matrix is assumed as $D_1 = [D(1), D(2), \dots, D(N_B)]$, and the m th row is represented as $D(m)$. By utilizing row permutation vector λ , after performing row sampling function on DCT matrix D_1 .

$$\psi_{chaos}^{DCT} = [D(\lambda(1)), D(\lambda(2)), \dots, D(\lambda(N_B))] \quad (6)$$

The produced chaotic DCT matrix is considered a scrambling matrix C which is a product of conventional DCT D_1 .

$$\psi_{chaos}^{DCT} = CD_1, \left(\psi_{chaos}^{DCT} \right)^t = D_1^t C^t \quad (7)$$

$$\psi_{chaos}^{DCT} \left(\psi_{chaos}^{DCT} \right)^t = CD_1 D_1^t = CJC^t = J \quad (8)$$

The term ψ_{chaos}^{DCT} is the chaotic DCT matrix.

Similarly, from the Chebyshev map, chaotic sequences are produced.

$$p(N + 1) = \cos(K \times ar \cos(p(N))) \quad (9)$$

Here, $p \in (-1, 1)$, the p_0 which is an initial value and K is considered as keys. Then various chaotic scrambling sequences B_1 are obtained with the aid of various control parameters and initial values. The obtained Z_{chaos}^{DCT} is assumed as the chaotic DCT measurement matrix result. To create a partial chaotic DCT matrix, select from N_B rows to M_B rows of the DCT matrix of size $N_B \times N_B$. It acts as a Θ_{chaos}^{DCT} MM of size $M_B \times N_B$.

$$\Theta_{chaos}^{DCT} = \sqrt{\frac{N_B}{M_B}} RMD_1 \quad (10)$$

where the subsampling matrix is denoted as R .

b. Key Generation Using SHA-256

The actual binary sequence is the SHA-256 function's input; after that, it produces a 256-bit key stream (BKS). Then divides the acquired 256 BKS *Key* into 8-bit blocks. The mathematical expression for *Key* is given below.

$$Key_1 = ke_1, ke_2, \dots, k_{32} \quad (11)$$

Based on the key stream of SHA-256, three respective values are generated that are mentioned as follows

$$v_1 = \frac{(ke_1 \oplus ke_2 \oplus \dots \oplus ke_{16})}{256} \quad (12)$$

$$v_2 = \frac{(ke_{17} \oplus ke_2 \oplus \dots \oplus ke_{32})}{256} \quad (13)$$

$$v_3 = \frac{(ke_1 \oplus ke_2 \oplus \dots \oplus ke_{32})}{256} \quad (14)$$

Here, XOR operation is defined by $\oplus$. The three initial values of the chaos model are derived in terms above values.

$$p_0 = p'_0 + v_1/2 \quad (15)$$

$$q_0 = q'_0 + v_2/2 \quad (16)$$

$$r_0 = r'_0 + v_3/2 \quad (17)$$

The chaos system's actual initial values are denoted as $p'_0 q'_0 r'_0$.

c. Optimal Key Selection Using Ebola Search Optimization Algorithm (ESOA)

In ESOA, every exposed individual position is updated by using the Eq. (18)

$$n(L_j^{t+1}) = nL_j^t + \alpha O(L) \quad (18)$$

where α denotes the individual displacement scale factor, the updated and the actual positions are denoted as nL_j^{t+1} and nL_j^t with respect to the time t and $t + 1$. The movement value produced by an individual is represented as $O(L)$.

$$O(L) = Srate * Rand(0, 1) + O(Inl_{best}) \quad (19)$$

$$O(S) = Lrate * Rand(0, 1) + O(Inl_{best}) \quad (20)$$

Exploitation Phase: It is constructed within a distance of '0'. The infected individual stays or is relocated inside a limit not beyond *Srate*. Here, the short distance movement is denoted as *Srate*.

Exploration Phase: It is established based on the infected individual migrating far off the average neighborhood range $Lrate$. From this, the conclusion is that the farther the movement, the more individuals present in A are exposed to infection. The neighborhood parameter is used to regulate the $Srate$ and $Lrate$. If the value is ≥ 0.5 , an individual moves after the neighborhood, which causes mega infection, or else an individual is present inside the neighborhood that curbs infection. Finally, it finds the optimal double secret key of a 256-bit key stream for security purposes.

Initialization. At first, the random number distribution is used to generate the initial population, for which '0' is the initial position for all individuals. By using Eq. (21), the individual is generated. For the i th position, the upper and the lower bounds are represented as u_i and l_i correspondingly. In the population size, the range L is from 1, 2, 3 ... n .

$$inl_i = l_i + Rand(0, 1) * (u_i + l_i) \quad (21)$$

The choice of the current best solution is calculated in the group of infected individuals at a time t .

$$bestA = \begin{cases} Gbest, & fitness(Cbest) < fitness(Gbest) \\ Cbest, & fitness(Cbest) \geq fitness(Gbest) \end{cases} \quad (22)$$

where, the best solution, current best solution, and global best solution are represented as $bestA$, $Cbest$ and $Gbest$ at a time t respectively. The OF implemented in the issue is mentioned as $fitness$.

Update the susceptible (S), infected (I), hospitalized (H), exposed (E), vaccinated (V), recovered (R), funeral (F), Quarantine (Q), and dead (D). In this research paper, the application of differential calculus's aim is to acquire the rate of change of S, I, H, R, V, D, and Q quantities with respect to time T .

$$\frac{\partial S(T)}{\partial T} = \mu - (\alpha_1 I + \alpha_3 D + \alpha_4 R + \alpha_2 (PE))S - (\lambda S + \Gamma I) \quad (23)$$

$$\frac{\partial I(T)}{\partial T} = (\alpha_1 I + \alpha_3 D + \alpha_4 R + \alpha_2 (PE)\tau)S - (\Gamma + \nu)I - (\lambda)S \quad (24)$$

$$\frac{\partial H(T)}{\partial T} = \beta I - (\nu + \omega)H \quad (25)$$

$$\frac{\partial R(T)}{\partial T} = \nu I - \Gamma R \quad (26)$$

$$\frac{\partial V(T)}{\partial T} = \nu I - (\pi + \nu)V \quad (27)$$

$$\frac{\partial D(T)}{\partial T} = (\lambda S + \Gamma I) - \zeta D \quad (28)$$

$$\frac{\partial Q(T)}{\partial T} = (\mu I - (\nu R + \Gamma D)) - \varepsilon Q \quad (29)$$

The expression mentioned above is scalar functions.

Objective Function. The PSNR value is considered the objective function in this GFDM-PON secure data transmission. The mathematical expression for objective function is given as follows

$$OBJECTIVE = Max(PSNR) \quad (30)$$

Updating Procedure for a New Key. After the fitness calculation, update the S, I, H, E, V, R, F, Q, and D. The new keys are updated to enhance the keys in the encryption and decryption phase. The mathematical expression for this new key updation can be written below

$$New_keyupdate = S + I + H + E + V + R + F + Q + D \quad (31)$$

d. **Symbol Scrambling Scheme to Modulate the CS Data**

An effective digital modulation symbol scrambling method is used to further improve the security of secure CS data. Consider $Z = [Z(0)Z(1) \cdots Z(n-1)]$ defining a BPSK symbol vector of size n , and the chaotic function is denoted as $O = [O(0)O(1) \cdots O(n-1)]$ of size n , it is obtained from the logistic map. The outcoming O sequence is converted into B binary sequence.

$$B(ke) = \text{sign}(O(ke) - 0.5) \quad (32)$$

Then $\Upsilon = [Y(0)Y(1) \cdots Y(n-1)]$ is the scrambled PBSK symbol vector, and it can be acquired as below

$$Y(ke) = Z(ke)B(ke)(L) \quad (33)$$

The inverse scrambling function is performed on the receiver side.

Secure CS Transmission Model. GFDM is a non-orthogonal waveform that transmits $n = mk$ modulation symbols for each block utilizing subsymbols of m and subcarriers of k . The mathematical model for transmitted GFDM symbol is expressed as follows:

$$p[N] = \sum_{k=0}^{k_1-1} \sum_{m=0}^{m_1-1} s_{k,m} g[\langle N - Mk_1 \rangle_n] e^{-j2\pi \frac{k}{k_1} N} \quad (34)$$

where $s_{k,m}$ denotes transmitted modulation symbols at m th subsymbol and k th subcarrier, the transmitted prototype filter is represented as $g[N]$, and the n operator modulo is $\langle (\cdot) \rangle_n$.

The matrix notation of GFDM is given by

$$p = Bs \quad (35)$$

where, $s = (s_0 \dots s_{m-1})^t$ and $s_M = (s_{0,M} \dots s_{k-1,M})^t$.

The above equation shows that the GFDM symbol is a linear modulation; the interaction among subsymbols and subcarriers is denoted as B .

The BPSK modulation is utilized to qualify and modulates the acquired CS binary sequence. Through the wireless channel, the modulated signal is transformed. After that, it maps the stream of bits into the symbol stream of BPSK, which is classified as sub-blocks. Every sub-block is represented as $Z = [Z(0)Z(1) \cdots Z(n-1)]$ of size n , it is scrambling based on the expression (33). The term Υ is the resultant encrypted signal vector, and with the aid of the IFFT unit, it is converted into a GFDM time domain signal.

$$y = F^H Y \quad (36)$$

where the discrete Fourier transform (DFT) matrix is represented as F from this, the element (m, o) can be derived as

$$F(m, o) = \frac{1}{\sqrt{n}} e^{-i \frac{2\pi MN}{n}} m, o \in \{0, 1, \dots, n-1\} \quad (37)$$

To eliminate the intersymbol interference, appends a cyclic prefix (CP) to the gained time domain GFDM signal $s = (s_0 \dots s_{m-1})^T$. Before transmitting the signal via the channel, the signal has to pass through the power amplification system.

At the receiver block, the noise of the received signal is,

$$R = H_1 \times s + W \quad (38)$$

where, the selective channel's discrete impulse response of the wireless frequency is denoted as, H_1 and additive white gaussian noise is denoted as $W = [\omega_0 \omega_1 \dots \omega_{n-1}]$. The quasi-static exponentially decaying multipath Rayleigh fading system is utilized in this research work. The mathematical model for his system is given below

$$h(t) = \sum_{i=0}^{i-1} h_i \zeta(t - iT_s) \quad (39)$$

where h_i is designed as a complex gaussian random variable.

$$h_j \sim cn(0, \theta_0 \exp(-iT_s/T_{RMS})) \quad (40)$$

Here $\theta_0 = 1 - \exp(-T_R/T_{RMS})$. The receiver sampling value is denoted as T_s , and the root mean square (RMS) is represented as T_{RMS} . At the k th subcarrier, the channel frequency response can be written as

$$H(k) = \sum_{i=0}^{i-1} h(i) \exp\left(-j \frac{2\pi ki}{n}\right), \quad (41)$$

where h_i represents the channel response at i th tap.

The FFT unit is used to process the outcoming signal R . Then, the data present at the k th subcarrier is expressed as

$$\Re(k) = H(k)Y(k) + G(k) \quad (42)$$

where the complex gaussian is denoted as $G(k)$ of zero mean and σ_k^2 variance in the frequency region.

Hence equalized the outcoming signal by minimum mean square error (MMSE) or zero-forcing (ZF). The expression for the equalized signal as

$$\check{Y}(k) = W(k)\Re(k) \quad (43)$$

For zero-forcing equalization $W(k) = \frac{1}{H(k)} = \frac{\check{H}^*(k)}{|\check{H}(k)|^2}$.

Equation (42) is substituted in the Eq. (43), then it produces

$$\begin{aligned} \check{Y}(k) &= W(k)[H(k)Y(k) + G(k)] \\ &= W(k)H(k)Y(k) + W(k)G(K) \\ &= Y(k) + G(k)/H(k) \end{aligned} \quad (44)$$

By using the chaotic scrambling sequence U , the equalized signal $\check{Y}(k)$ get descrambled. The expression for the acquired decrypted signal $\check{S}$ is

$$\begin{aligned} \check{S}(k) &= \check{Y}(k)U(k) \\ &= [Y(k) + G(k)/H(k)]U(k) \\ &= [S(k)U(k) + G(k)/H(k)]U(k) \\ &= S(k) + G(k)U(k)/H(k) \end{aligned} \quad (45)$$

The obtained decrypted BPSK symbol stream is converted into a bit stream and processed by inverse quantization. The derived data is utilized to create a sub-block of sequence. Then, the entire sub-block sequence is integrated to produce the actual binary sequence.

3 Results and Discussion

This section explains the proposed method's results to prove the proposed model's effectiveness when compared with other existing techniques. This proposed method has been implemented in a MATLAB platform. The performance matrices such as PSNR, entropy, CCDF, encryption time, decryption time, BER, and MSE are analyzed and compared with existing techniques.

A. Performance metrics

The two factors used to estimate the recovered data quality are PSNR and MSE. We provide the formula below.

$$\mathbf{MSE.} \quad MSE = \frac{\sum_1^n [p - q]^2}{m * n} \quad (46)$$

where p denotes the actual signal, q denotes the recovered signal, and the number of rows and columns of the input signal is denoted as m and n correspondingly.

$$\text{PSNR. } PSNR = 10 * \log_{10} \frac{M^2}{MSE} \quad (47)$$

where M represents the maximum value of the signal.

$$\text{Entropy. } H(S) = \sum_{i=0}^{l-1} p_i \log_2 \frac{1}{p_i} \quad (48)$$

where p_i represents the probability of the symbol s_i occurring S .

Bit Compression Ratio. The bit compression ability of the proposed system is measured by using compression ratio. The below expression is used to compute the BCR.

$$\lambda = \frac{CR \times N_{ai}}{N_{actual_bit}} \times 100\% \quad (49)$$

where the actual transmission bits are denoted as n_{actual_bit} , and the required transmission bit is denoted as N_{ai} .

Figure 2 illustrates the CCDF of PAPR of some existing and proposed models. The conventional OFDM signal and the secure OFDM signal are the two existing systems compared with the proposed GFDm system. Based on the varying CR values of 0.25, 0.5, and 0.75, the deviation of the signals may increase rapidly. When the PAPR value is 4, the attained values of the original signal are at 1.004402, and the obtained CCDF values are 1.004402, 0.998001, and 0.987036. If the value of PAPR is 8, the attained values of original and based on varying CR values are 1.005161, 1.005161, 1.000994, and 0.987852, respectively.

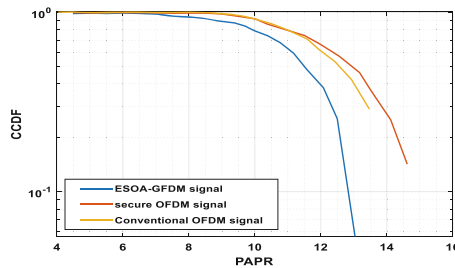


Fig. 2. Comparison of CCDF values of proposed with existing methods

Figure 3 demonstrates the BER performance of the introduced signal with varying compression ratios (CR) like 0.75, 0.5, and 0.25. When the received optical power is -24 dBm, the attained BER values for original CR is at 0.75, 0.5, and 0.25 are 0.246, 0.278, 0.174 and 0.217 respectively. If the received optical power is at -14 , the achieved BER values are 1.38×10^{-8} , 1.59×10^{-8} , and 1.71×10^{-8} for the original, and the CR values

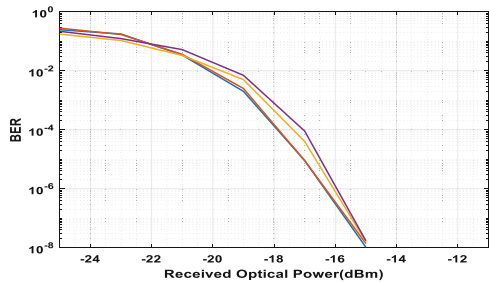


Fig. 3. BER values by varying CRs

of about 0.75, 0.5, and 0.25. Figure 4 clearly shows that the proposed method achieves significantly low BER values.

Table 1 represents the ET and the decryption time of the proposed and the existing systems like chaotic compressive sensing in OFDM-PON, chaos and DNA encoding in OFDM-PON, Brownian motion, and chaos in OFDM-PON. If 524288 bits are considered an amount of data, the ET and the DT of the proposed system are 0.054 and 0.048. The obtained ET of existing chaotic compressive sensing in OFDM-PON is 0.078, implemented Chaos and DNA encoding in OFDM-PON is 0.30, and previous Brownian motion and chaos in OFDM-PON are 0.99. The decryption time of existing chaotic compressive sensing in OFDM-PON is 0.15, and implemented chaos and DNA encoding in OFDM-PON is 0.30. Previous Brownian motion and chaos in OFDM-PON are 0.92. The attained ET and DT of existing methods are maximum when compared with the proposed model, so the GFDM-PON system is more efficient than the other systems.

Table 1. Comparison of et and decryption time.

	Proposed algorithm in GFDM-PON	Chaotic compressive sensing in OFDM-PON [17]	Chaos and DNA encoding in OFDM-PON [19]	Brownian motion and chaos in OFDM-PON [18]
Amount of data (bits)	524288	524288	524288	524288
Encryption time (s)	0.054	0.078	0.30	0.99
Decryption time (s)	0.048	0.15	0.30	0.92

Table 2 demonstrates PSNR, MSE, and entropy performance comparison of proposed EOSA with some of the existing Spark optimization algorithm (SOA), Hive optimization algorithm (HOA), and Particle swarm optimization (PSO). It mentions that the proposed EOSA achieves better PSNR, MSE, and entropy values than the other three existing

techniques. The achieved PSNR values of proposed EOSA and existing SOA, HOA and PSO are 57.86319 and 56.76692, 55.78557, and 55.32411, correspondingly. Maximum PSNR values produce high quality data as the resulting outcome; thus, the proposed model achieves high PSNR value using the EOSA approach. The accomplished MSE values of existing SOA are 0.008166, HOA is 0.012772, and PSO is 0.015325. Compared with the proposed EOSA, the obtained value is 0.004114, which is the minimum. It proves that the proposed algorithm achieves better data accuracy in the receiver block. The entropy value of the proposed method is 9.495201, and the existing SOA, HOA, and PSO are 9.411404, 9.401598, and 9.401598. The encryption and decryption process is performed effectively due to the maximum entropy value of the proposed system.

Table 2. Performance metric comparison of proposed and existing methods

Performance metrics	Proposed EOSA	SOA [20]	HOA [21]	PSO [22]
PSNR	57.86319	56.76692	55.78557	55.32411
MSE	0.004114	0.008166	0.012772	0.015325
Entropy	9.495201	9.411404	9.401598	9.401598
Encryption time	0.054	0.062	0.074	0.089
Decryption time	0.048	0.059	0.073	0.079

Figure 4 (a) and (b) demonstrate the ET and DT, introduced with other optimization techniques. ET of the proposed EOSA is minimum, i.e. 0.054 when compared with the existing SOA, HOA and PSO are 0.062, 0.07,4, and 0.089, accordingly. Also, the time taken for decrypting the data is 0.048, which is comparatively less compared to the already implemented optimization algorithms and its attained values are 0.059, 0.07,3, and 0.079. Hence our proposed EOSA approach is considered an effective optimization algorithm among all the compared techniques.

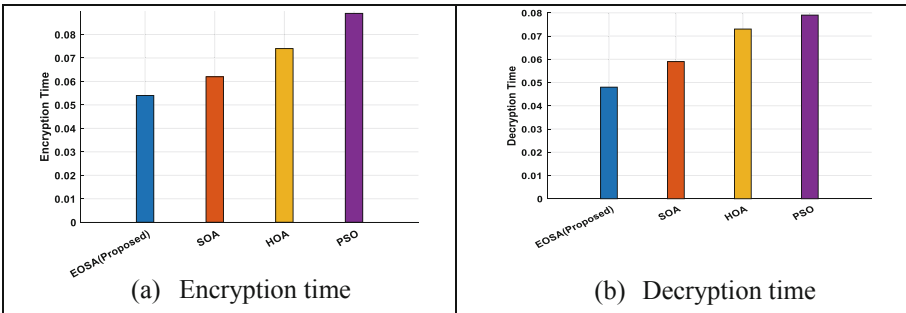


Fig. 4. Comparison of (a) ET (b) DT with existing methods

Figure 5 compares the proposed fitness function with previous optimization algorithms. The proposed EOSA technique achieves maximum fitness function from a minimum number of iterations. But the existing SOA, HOA, and PSO require more iteration to produce better fitness functions. It leads to face-time complexity during the key selection process. But the proposed algorithm's time consumption is significantly less and produces the highest fitness value than other optimization algorithms.

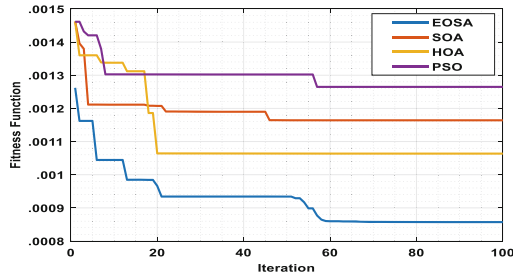


Fig. 5. Comparison of fitness function values

The distribution of the measurement matrix with the first row, first column, and the entire matrix is shown in Fig. 6. The obtained quality of the proposed EOSA-GFDM signal is maximum, with different value ranging from 0–1. If the range is at 0.6, the proposed method achieves the peak value as 50.

In the first row. In the first column, the quality of the signal is maximum in the range of 0.4. Finally, in the entire matrix, the proposed EOSA-GFDM signal accomplishes peak value at the range of 0.5, and the quality of the produced signal is 6000. The graph below concludes that the proposed EOSA.

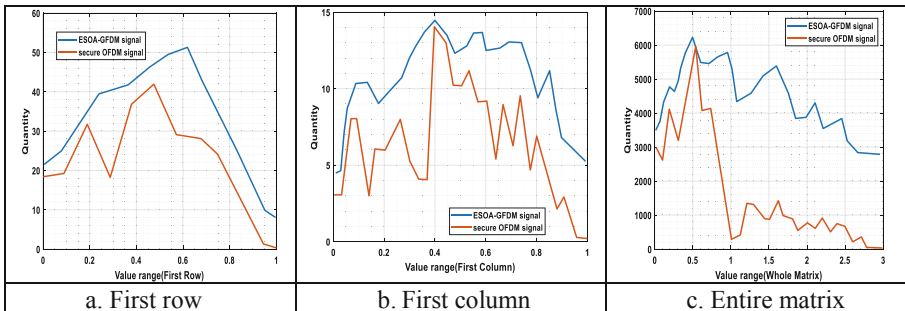


Fig. 6. The distribution of measurement matrix with (a) First row (b) First Column (c) Entire matrix

4 Conclusion

This is the first research work that applies CS methods to the data transmission for GFDM-PON in the physical layer security. The major aim of this proposed technique

is to compress and transmits the data in the same bandwidth. Moreover, to improve the security of the proposed model by the joint secure compressive sensing encryption technique and symbol scrambling strategy. Two chaotic DCT matrixes act as the sparse basis and measurement matrix for the block-based CS data. Moreover, the proposed system attains a better PSNR value and very low computational cost. After that, the generated BPSK signal from the CS block is further encrypted with a chaotic sequence transmitted into a wireless GFDM system. The achieved BER value of the introduced model is comparatively low. To improve PLS in the GFDM-PON, a key generation and optimal key selection have been performed. The Ebola meta-heuristic strategy is the optimization algorithm used in this research to choose the optimal key. At last, the joint secure compressive sensing encryption and symbol scrambling strategy improves the encrypted CS data's security in the GFDM-PON system. This proposed system will concentrate more on data encryption and decryption in future work.

References

1. Sánchez, J.D.V., Urquiza-Aguiar, L., Paredes, M.C.P., Osorio, D.P.M.: Survey on physical layer security for 5G wireless networks. *Ann. Telecommun.* **76**(3–4), 155–174 (2020). <https://doi.org/10.1007/s12243-020-00799-8>
2. Nandan, N., Majhi, S., Wu, H.C.: Beamforming and power optimization for physical layer security of MIMO-NOMA based CRN over imperfect CSI. *IEEE Trans. Veh. Technol.* **70**(6), 5990–6001 (2021)
3. Sultan, A., Yang, X., Hajomer, A.A.E., Hu, W.: Chaotic constellation mapping for physical-layer data encryption in OFDM-PON. *IEEE Photonics Technol. Lett.* **30**(4), 339–342 (2018)
4. Katsuki, Y., de Abreu, G.T.F., Ishibashi, K., Ishikawa, N.: Noncoherent massive MIMO with embedded one-way function physical layer security, p. 11893 (2022)
5. Qi, Q., Chen, X., Zhong, C., Zhang, Z.: Physical layer security for massive access in cellular internet of things. *Sci. China Inf. Sci.* **63**(2), 1–12 (2020)
6. Ghanbarisabagh, M., Vetharatnam, G., Giacomidis, E., Rouzegar, H., Mallouki, N.: A Survey on High-Capacity OFDM-based Passive Optical Networks. 4692 (2018)
7. Maeng, S.J., Yapici, Y., Guvenc, I., Bhuyan, A., Dai, H.: Precoder design for physical-layer security and authentication in massive MIMO UAV communications. *IEEE Trans. Veh. Technol.* **71**(3), 2949–2964 (2022)
8. Wang, W.Q., Zheng, Z.: Hybrid MIMO and phased-array directional modulation for physical layer security in mmwave wireless communications. *IEEE J. Sel. Areas Commun.* **36**(7), 1383–1396 (2018)
9. Wu, Y., Khisti, A., Xiao, C., Caire, G., Wong, K.K., Gao, X.: A survey of physical layer security techniques for 5G wireless networks and challenges ahead. *IEEE J. Sel. Areas Commun.* **36**(4), 679–695 (2018)
10. Zhu, F., Gao, F., Lin, H., Jin, S., Zhao, J., Qian, G.: Robust beamforming for physical layer security in BDMA massive MIMO. *IEEE J. Sel. Areas Commun.* **36**(4), 775–787 (2018)
11. Zhang, C., Zhang, W., He, X., Chen, C., Zhang, H., Qiu, K.: Physically secured optical OFDM-PON by employing chaotic pseudo-random RF subcarriers. *IEEE Photonics J.* **9**(5), 1–8 (2017)
12. Hu, Z., Chan, C.K.: A real-valued chaotic orthogonal matrix transform-based encryption for OFDM-PON. *IEEE Photonics Technol. Lett.* **30**(16), 1455–1458 (2018)
13. Bi, M., et al.: Chaotic arnold transform and chirp matrix encryption scheme for enhancing the performance and security of OFDM-PON. *Opt. Fiber Technol.* **51**, 64–70 (2019)

14. Li, M., et al.: Flexible non-linear physical security coding scheme combined with chaotic neural network for OFDM-WDM-PON. *Opt. Express* **29**(16), 25848–25858 (2021)
15. Fu, X., et al.: A chaotic modified-DFT encryption scheme for physical layer security and PAPR reduction in OFDM-PON. *Opt. Fiber Technol.* **42**, 126–131 (2018)
16. Deng, L., et al.: Secure OFDM-PON system based on chaos and fractional fourier transform techniques. *J. Lightwave Technol.* **32**(15), 2629–2635 (2014)
17. Wu, T., et al.: Compressive sensing chaotic encryption algorithms for OFDM-PON data transmission. *Opt. Express* **29**(3), 3669–3684 (2021)
18. Wu, T., et al.: Security enhancement for OFDM-PON using brownian motion and chaos in cell. *Opt. Express* **26**(18), 22857–22865 (2018)
19. Zhang, C., Zhang, W., Chen, C., He, X., Qiu, K.: Physical-enhanced secure strategy for OFDMA-PON using chaos and deoxyribonucleic acid encoding. *J. Lightwave Technol.* **36**(9), 1706–1712 (2018)
20. Yu, H.: Apriori algorithm optimization based on spark platform under big data. *Microprocess. Microsyst.* **80**, 103528 (2021)
21. Liu, Y., Chang, X., Zhou, Y., Dang, S., Shi, J., Liu, Z.: HIVE: route optimization problem of automatic delivery system with mobile depot and multiple drones. In: 2021 IEEE 23rd International Conference on High Performance Computing & Communications; 7th International Conference on Data Science & Systems; 19th International Conference on Smart City; 7th International Conference on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)), pp. 1338–1345 (2021)
22. Wang, F., Zhang, H., Zhou, A.: A particle swarm optimization algorithm for mixed-variable optimization problems. *Swarm Evol. Comput.* **60**, 100808 (2021)

Software-Defined Network and Telecommunication



CPC-SDN: A Centralized Proactive Caching Based on Software Defined Network

Sembati Yassine¹, Naja Najib¹(✉), and Jamali Abdellah²(✉)

¹ Institut national des postes et télécommunications, Rabat, Morocco

naja@inpt.ac.ma

² Ecole national des sciences appliquée, Berrechid, Morocco

jamali.uh1@gmail.com

Abstract. The ongoing development and evolution of the Internet highlighted the limitations of its architecture and the need for a new design. Simultaneously, as a result of the development of new applications, new requirements emerged, signaling to the research community that fundamental modifications to the Internet paradigm were required. In this scenario, NDN (Named Data Networks) and SDN (Software-Defined Networks) stand out as two of the key proposals for the future Internet. Nonetheless, NDN's effective data transport and SDN's adaptable management structure look to function well together to address the current problems with Internet design.

In this paper, we introduce Centralized Proactive caching based on Software defined network "CPC-SDN", a novel NDN-based cache solution controlled by an OpenFlow controller. CPC-SDN allows for clean state integration between NDN and SDN as well as proactive caching techniques depend on the frequency of contents that are entirely controlled by the controller.

Keywords: Named data network · Routing · Software defined network · Proactive cache · OpenFlow

1 Introduction

The Internet has experienced phenomenal growth and is becoming more and more prevalent in people's daily lives. The accelerated growth of the Internet and its rising popularity among people have forced a reevaluation of the traditional idea of its architecture and the creation of new paradigms. NDN and SDN technologies are among the most promising concepts for a new architecture. The present Internet is based on the Internet Protocol (IP) and takes an end-to-end communication paradigm into account, this concept addresses the first Internet user requirements, which are confined to fundamental applications such as video and file sharing. The growth of these applications has drastically changed how Internet traffic behaves. The Internet architecture can evolve from a straightforward destination-forwarding paradigm to a more sophisticated architecture concentrating on content data and users instead of hosts to address the growing use of information distribution instead of information exchange between end nodes. A named data network is the architecture of these concepts [1, 2].

There are two sorts of communications on the NDN: interest and data. A user sends content requests using the interest message. A Data Message is sent by the content producer to the user requester in response to his content request. A cache replacement strategy is not supported by the existing architecture. Currently, NDN incorporates several techniques, including Caching Everything Everywhere (CEC) [3] and Least Recently Used (LRU) [3]. Those caching solutions are incapable of producing the results required for NDN architecture. As a result, a new replacement method must be proposed to improve the performance of the NDN architecture.

To demonstrate the viability of NDN, the Software-Defined Networking (SDN) paradigm [4] could be used. SDN enables the integration of new caching and forwarding systems using current hardware components, as it allows network components to be highly programmable. Additionally, because SDN enables network virtualization, it reduces material costs and complexity while increasing network functionality. The standard for implementing SDN communication with the underlying network is typically OpenFlow [6].

We plan to leverage the SDN architecture in this study to improve the management performance of caches in the NDN network, hence we propose a unique technique called CPC-SDN that integrates the SDN and NDN paradigms via a clean state integration based on NDN flows managed by an OpenFlow controller. CPC-SDN optimizes network resources while also enabling flexibility, scalability, and ease of configuration. Two key contributions are made by our proposed solution:

(i) By adjusting the underlying protocol, OpenFlow switches can successfully support NDN functions in the network as routing and caching; (ii) support for programmable forwarding by name and proactive caching in NDN.

The implementation of a new proactive, centralized SDN-based architecture for caching that enables both the storage of the most popular content as well as the processing and pushing off that popular content based on the volume of interest expressed in each content to the switch's content store that is closest to the user.

The rest of this article is structured as follows. Section 2 introduces the Named Data Network. The notion of SDN is introduced in Sect. 3. Section 4 then demonstrates the relevant work. Section 5 describes our methodology. Section 6 shows the results of our experiments. Section 7 ends the report and discusses future research.

2 Named Data Networking

The NDN architecture [1] was first proposed in 2009 by Van Jacobson and the NDN project. The NDN solution is built around naming content. The name material in NDN is organized hierarchically. A hierarchical name is organized like 'NDN:/inpt.ma/video1/version1/' with a prefix and suffix. When a user wishes to obtain material in the NDN architecture, he sends an interest message to the network, which is responded to by a data message containing the data. An interest message in the network will be responded to by a data message containing the data. Interest messages include the name of the data, i.e., the data's prefix, as well as a nonce and many optional option fields, such as the request's validity term or the public key of the owner who signed the message. Data messages are made up of the data's name (the same prefix as the

interest message), the data's content, and the content's signature. These messages leave the network with the essential responsibility of resolving the names of the requests and enabling data routing to the user and message routing in the network.

Since the NDN architecture is itself a network of caches, several copies of the same content can be present at different network nodes. The data can thus come from different points in the network and no data is associated a priori with a location in the network. Each NDN consists of three main components shown in Fig. 1 the Content Store (CS), the Pending Interest Table (PIT), and the Forwarding Information Base (FIB). It uses a face notion in its architecture that is comparable to the IP interface. It comprises both physical connections and application interfaces.

- The FIB keeps track of the inputs where interest messages are to be transmitted to potential sources of content.
- The PIT allows a node to maintain a list of the interest messages it has transmitted and the interfaces on which they were received, to avoid unnecessary retransmissions. The PIT entries are also used so that Data messages can follow the reverse path of Interest messages. Finally, the CS is a cache that stores local copies of the content. Each NDN node will thus be able to respond directly to requests (Interest) if it has the contents in its Content Store.

Since the network nodes can keep a local copy of the content that has passed through them, the content will therefore be disseminated in the network and be closer to the users. Thus, the interests will potentially encounter a closer node with the content and not be propagated back to the originating server, which will reduce the load on that server and limit the impact of a search in the network. This will also reduce content access times

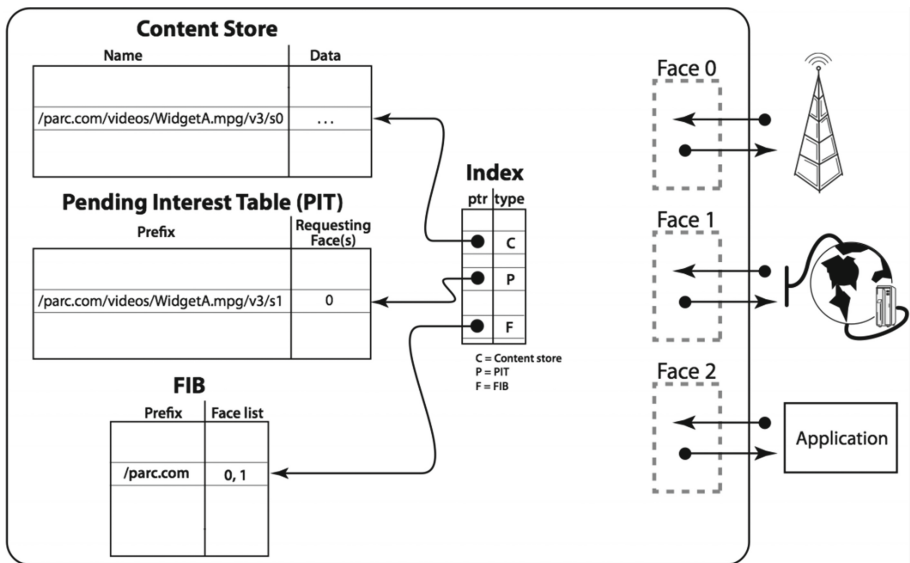


Fig. 1. NDN architecture [1]

for users and improve their quality of experience (delay, throughput). Since the nodes' cache memory is constrained, numerous strategies have been proposed to choose which information to store in memory and which to delete when the memory is full in the context of managing the cache memory of nodes [7, 8].

3 Software Defined Network

In this part, we will first discuss SDN [5, 6], a new method for increasing network programmability that has resulted in the development of new standard communication protocols such as OpenFlow [7, 8]. We also go through the principles of the OpenFlow standard and the OpenFlow protocol's characteristics.

3.1 SDN Architecture

The SDN architecture is conceptually composed of three layers or planes: data, control, and applications (see Fig. 2). Communication between each plane occurs through programmable interfaces (APIs, which stand for Application Programming Interface).

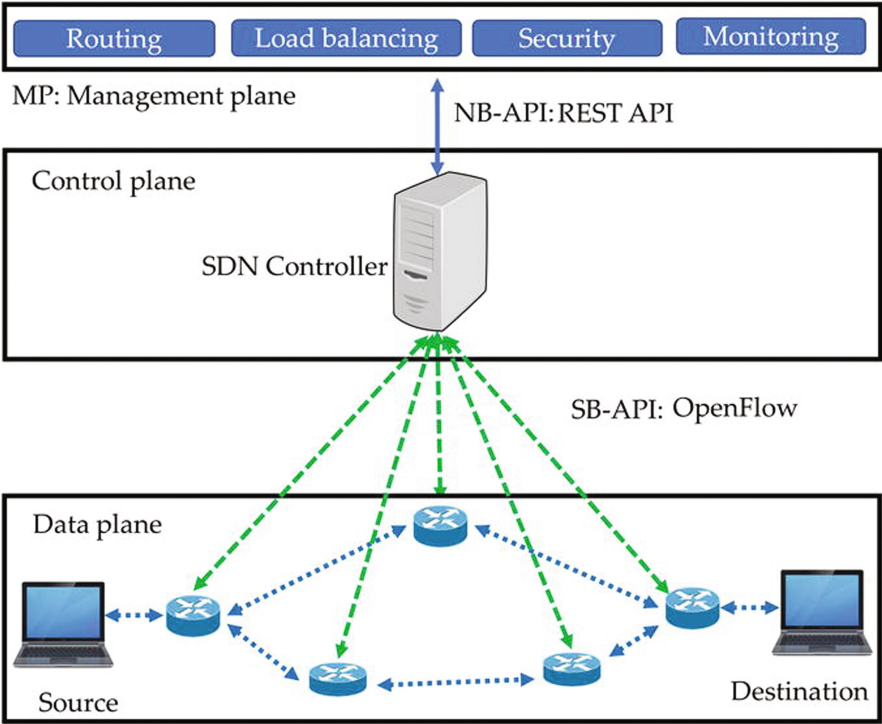


Fig. 2. SDN architecture [25].

In SDN the data plane is found in each of the network devices (as in the current network model), but the control plane is only present in the controller. The following is a detailed analysis of each of the planes to better understand how they work.

Data Plane

It is composed of all network devices (such as switches, routers, and access points) that are exclusively responsible for the physical routing of packets. This plane's elements can be either tangible (hardware) or virtual (software). Their behavior or operation as a network element is determined by the SDN controller's rules. The SDN infrastructure does not make judgments about packet trajectory this is handled by the control plane.

Control Plane

It is responsible for centralizing the management of all information flow through the data layer. It permits the implementation of all switching and packet routing policies through a controller, eliminating the requirement to do so in each piece of interconnection hardware in the network [9].

In this plane, the controller placement may be either distributed or centralized. Small-scale networks frequently employ centralized control, where a single controller is in charge of overseeing all network components. The single point of failure for the entire control system and its lack of scalability are its weaknesses.

Application Plane

This plane contains various applications that provide several functionalities that determine the behavior and characteristics of the architecture. The northbound APIs allow the communication and interaction of these applications with the control layer and can be executed both outside and inside the controller software. Among the main categories or groups into which the applications in this plane can be divided are traffic engineering, measurement, monitoring, safety, reliability, and management tools.

The exchange between the data plane and the control plane is done via the southbound APIs, with OpenFlow as the standard protocol. Through this API the controller can send a set of policies or configurations, control the scheduling of all forwarding operations, perform statistics reporting, and send/receive event notifications from the data plane devices.

3.2 OpenFlow Protocol

It is the only standardized SDN protocol. It is used by default by the southbound APIs for communication between the control plane and the data plane devices.

Through OpenFlow, the controller establishes the connection with the data layer devices, both physical and virtual, without the need to use routing protocols. By programming the rules for each flow, it is possible to obtain granular control of the network, thus achieving a quick response to changes required by services and applications. For an exchange with the controller, the devices in the data layer must support the OpenFlow protocol. These devices are divided into two categories: pure OpenFlow or hybrid.

4 Related Works

While the concepts of Named Data Networks and Software Defined Networks have been studied separately in the past, combining the two has become a new research topic. While SDN can leverage the potential of NDN caching while monitoring and controlling the whole NDN network, it can also be used as an example to monetize the network potential of NDN.

Numerous cache replacement ideas have been put forth to enhance cache performance in the NDN network [10–12]. Distributed and centralized strategies make up the two categories. While the former is primarily based on the NDN architecture, the latter is based on NDN cache replacement methods based on SDN [12, 13].

The term “universal caching” (UC) refers to a cache replacement policy that was established in [14, 15]. The strategy was assessed and contrasted with FIFO and LRU. The concept outperforms others, according to the authors, even when the CS size and topology change.

A replacement policy based on content popularity was suggested by the study [12], called content caching popularity (CCP). The authors’ research was based on a network with consumers and producers situated at its periphery. They assessed and contrasted the cache hit, server load, and average packet transfer speed performance of LRU and LFU. Jmal and colleagues developed a centralized SDN architecture for NDN replacement management. In their approach, the controller controls all aspects of popularity estimation and cache storage decision-making. [16–18, 23] proposed using the SDN architecture for NDN. These systems improved on the NDN approach in various ways, including IP integration, routing scalability, the use of innovative transport technologies such as the routing protocol, and connectivity with the SDN architecture via OpenFlow. These studies, however, do not contain a solution for cache methods.

5 CPC SDN

The main problem with combining SDN OpenFlow with NDN in a hybrid architecture is that NDN cannot be handled in the switching field IP address of OpenFlow packets by existing OpenFlow IP-based standards. Studies in the literature that attempted to connect OpenFlow and NDN either proposed an overlay solution in which NDN packets are encapsulated in traditional IP packets and sent over the IP network or suggested a direct connection. The OpenFlow switches can then process these packets as they would any other IP packets. However, this approach has limitations, such as increased overhead and reduced efficiency. Another study proposes adding additional modules or plugins to OpenFlow to manage NDN processes [19, 20]. There are a few considerations for getting OpenFlow to a clean state for NDN integration. In NDN caching, each intermediary node between the source and the destination stores the requested content. This standard policy, called “Leave Copy Everywhere” (LCE), demonstrates significant redundancy and ineffective resource utilization in the caching system. The two primary components of CPC-SDN are:

1. Integrating SDN and NDN by providing NDN functionality to the core OpenFlow components This research aims to build an SDN-based control plane capable of controlling NDN packets using an OpenFlow controller called POX [21].
2. NDN-managed centralized proactive caching. The OpenFlow controller can keep the popular content of switches and users in the cache of the system on which it is running, increasing the availability of data. The CPC SDN architecture, depicted in Fig. 3, consists of NDN hosts that may act as both producers and clients, as well as an OpenFlow controller with a caching capability.

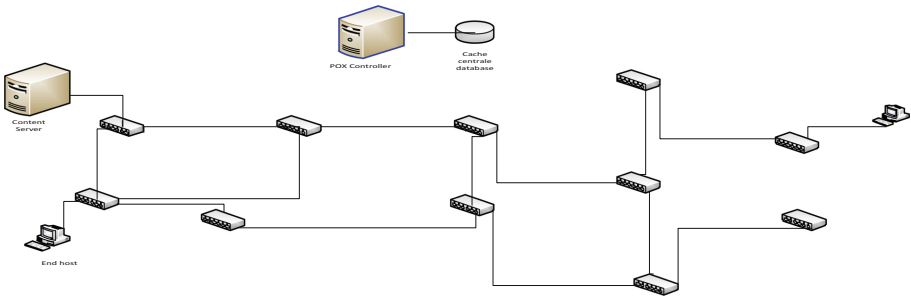


Fig. 3. CPC-SDN Architecture

5.1 OpenFlow-NDN Controller

The controller is the entity in charge of transmitting datapath architecture choices. The OpenFlow controller and the OpenFlow switches are critical SDN innovations.

One of the primary purposes of NDN is to allow caching in the network, which will increase content availability for destination requesters. Although switches are not designed to cache specific data to improve network availability, we suggest that the SDN-NDN controller utilize the cache of the system on which it operates to hold the content provided by producers and switch providers through content announcement messages. The necessary information can then be retrieved directly from the controller instead of contacting the content provider.

Whenever a switch transmits an interest message to a controller, the latter matches its internal cache to notice if the desired data is still available. If the requested content is already in the controller's cache, it can immediately send it to the switch, which will respond to the request as soon as possible (Algorithm 1). When compared to the situation where the switch must connect to another cache or the content provider to obtain the content, it can decrease round-trip time and improve the cache-hit ratio. The controller can also "proactively" send content to the switches when it predicts a spike in demand for a particularly popular piece of content shortly by using the priority field embedded into the controller's cache database. When the controller receives well-liked content, it gives it top priority and caches it. The data with the greatest priority is pushed to the

content storage of switches by a controller algorithm that routinely assesses the priority of the information in the cache (Algorithm 2).

Algorithm 1 CPC-SDN topology management

Entry: Switch S; Host H; controller SDN;/
H: sends announcement message to his connected S
S: Store the content and send announcement to the controller
S: Built the local cache
SDN: Built the centralized SDN cache with the content announcement
End.

Algorithm 2 CPC-SDN cache management.

Entry: Switch S; Host H; controller SDN;
H: sends Interest message to S;
S: receive the interest and checks the (CS, PIT, FIB);
IF (request found retour data to the host)
Else
S: Send interest to SDN
SDN: Receive interest and checks the global CS;
SDN: Push data to CS S proactively for future request
SDN: Increment the priority field of content
End.
End.

5.2 Proactive Caching

The term “proactive caching” refers to a mechanism that allows the controller to prefetch and store data in the switches’ caches. This can be used to ensure that popular content

is close to the user. We propose installing a proactive caching module in the OpenFlow-Named Data Network controller to store frequently used objects. The switches receive popular items from the controller's cache using the cache's "priority" field. We have assigned a priority of "0" to this material in the controller cache. The proactive caching mechanism in the controller frequently analyzes the relevance of the cached data.

We confirm that the proposed caching mechanism identifies this content data in the cache and informs the controller to deliver the material to all associated switches for storage in their caches. The switches' content stores are verified to ensure that the priority data is saved in them.

6 Experiments and Results

In the Ubuntu 14.04 environment, we tested the CPC-SDN architecture using an Abilene topology with one controller and 11 OpenFlow switches, which have been updated to support NDN functionalities, as well as two hosts: one as a client and the other as a content producer. The CPC-SDN architecture is centered around a POX controller, with all switches logically attached to it and two switches connected to a single node. The POX [15] controller was chosen for its adaptability and development freedom. We have added a new POX component called "CPC" that handles NDN networking management and forwarding. Our newly created module already includes the planned centralized Proactive Cache solution.

OpenFlow messages are used to implement the CPC-SDN functionalities. Data and interest packets are reported to the controller in OF Packet-In messages from OpenFlow switches. Data packets are sent using OF Packet Out, and switches notify the POX controller of their content via the OFP announcement message. The controller proactively pushes popular content to the switches' CS via OF ADD CS messages.

We compare CPC-SDN to flooding, the default method used in NDN on the Abilene topology. Our assessment is based on two key metrics: the number of network messages sent and the performance of network caches (cache hits). To begin with, because no memory is deployed in the network in the 0% case, the cache hit is always zero. The Cache Hit Ratio then increases in the 50% case before falling below this level in the 100% scenario because two nearby nodes will have the same cached information with a few minor differences. As the distance between caches increases, the likelihood of discovering a diversity of contents also increases. Depending on the situation, we observe that the Cache Hit Ratio improves with the amount of content stored for CPC and flooding (Fig. 4).

CPC-SDN achieves up to 60% Cache Hit performance for the 50% scenario, which means that every second request is resolved in the network, which is three times more efficient than the flooding technique. In general, CPC-SDN performs far better than flooding in terms of the Cache Hit Ratio.

Figure 5 shows that, when compared to the flooding used by default in NDN, our CPC-SDN technique reduces the number of Interest messages delivered in the network by 60% in the case where no cache is deployed in the network nodes (red and black curves), owing to SDN. Because all nodes are equipped with a cache, the difference in the number of overhead messages between CPC-SDN and Flooding has reduced in the scenarios with 100% cache, resulting in fewer messages being sent in the network.

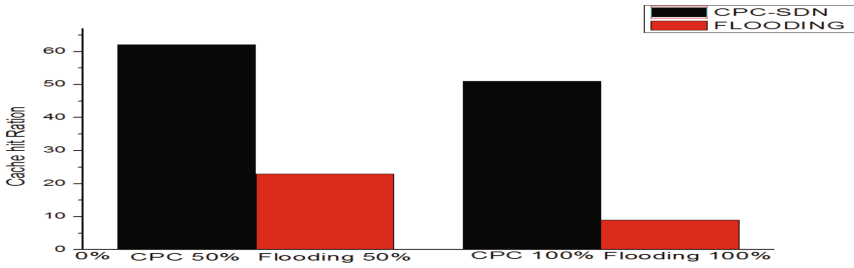


Fig. 4. Evaluation cache hit CPC-SDN compared to default NDN strategy

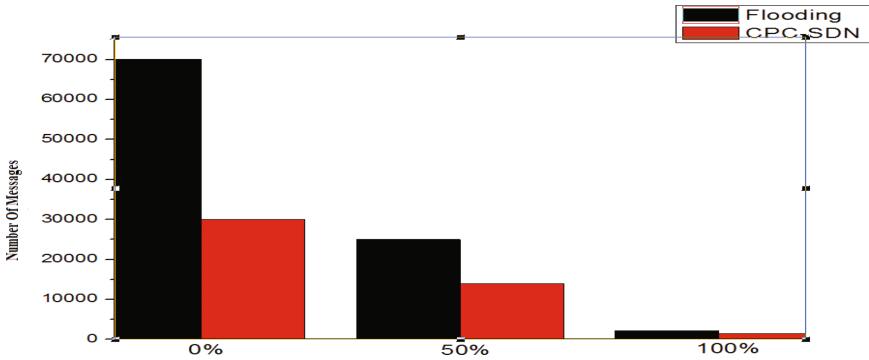


Fig. 5. Evaluation of overhead message CPC-SDN compared to default NDN strategy

7 Conclusion and Future Perspectives

The future of the internet could benefit from the combination of SDN and NDN. To achieve this, we tested the CPC-SDN approach, which gives SDN controllers control over the NDN cache. With the aid of NDN features like caching, we demonstrated how CPC-SDN permits programmable forwarding and allows us to deploy proactive caching mechanisms based on the number of request demands for content. Using an OpenFlow POX controller and OpenFlow switches, we put our solution into practice in a real environment. The results of the experiments demonstrated the beneficial impact of employing the SDN concept on the NDN network. This integration increased the percentage of cache hits while decreasing the number of overhead messages, resulting in improved NDN network performance.

Future work will involve expanding our architecture to test the network's scalability as well as enhancing other NDN features like routing by name and resource management. We also intend to compare our solution to other implementations of centralized caching.

References

1. Saxena, D., Raychoudhury, V., Suri, N., Becker, C., Cao, J.: Named data networking: a survey. *Comput. Sci. Rev.* **19**, 15–55 (2016). <https://doi.org/10.1016/j.cosrev.2016.01.001>

2. Zhang, L., et al.: Named data networking. *ACM Comput. Commun. Rev.* (2014)
3. Bernardini, C., Silverston, T., Vasilakos, A.: Caching strategies for information centric networking: opportunities and challenges. *CoRR*, pp. 54–61 (2016)
4. Schaller, S., Hood, D.: Software defined networking architecture standardization. *Comput. Stand. Interfaces* **54**, 197–202 (2017)
5. Kreutz, D., Ramos, F.M.V., Verissimo, P.E., Rothenberg, C.E., Azodolmolky, S., Uhlig, S.: Software-defined networking: a comprehensive survey. *Proc. IEEE* **103**(1), 14–76 (2015). <https://doi.org/10.1109/JPROC.2014.2371999>
6. Li, W., Meng, W., Kwok, L.F.: A survey on OpenFlow-based software defined networks: security challenges and countermeasures. *J. Netw. Comput. Appl.* **68**, 126–139 (2016). ISSN 1084-8045
7. Chai, W.K., He, D., Psaras, I., Pavlou, G.: Cache “less for more” in information-centric networks. In: Bestak, R., Kencl, L., Li, L.E., Widmer, J., Yin, H. (eds.) *NETWORKING 2012*. LNCS, vol. 7289, pp. 27–40. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30045-5_3
8. Bernardini, C., Silverston, T., Festor, O.: Socially-aware caching strategy for content centric networking. In: 2014 IFIP Networking Conference, pp. 1–9 (2014)
9. Chanda, A., Westphal, C.: A content management layer for software-defined information centric networks. In: *Proceedings of the 3rd ACM SIGCOMM Workshop on Information-Centric Networking, ICN 2013*, pp. 47–48. ACM, New York (2013)
10. Abidi, A., Gammar, S.M.: Towards new caching strategy for information-centric networking based on data proximity control. In: *IEEE International Conference on Computer and Information Technology*, pp. 540–547 (2015)
11. Zhu, Y., Mi, Z., Wang, W.: Cache probability replacement policy based on content popularity in content centric networks. *J. Electron. Inf. Technol.* **35**(6), 1305–1310 (2014)
12. Jmal, R., Fourati, L.C.: An OpenFlow architecture for managing content-centric-network (OFAM-CCN) based on popularity caching strategy. *Comput. Stand. Interfaces* **51**, 22–29 (2017)
13. Kalghoum, A., Gammar, S.M.: Towards new information centric networking strategy based on software defined networking. In: *Wireless Communications and Networking Conference*, pp. 1–6 (2017)
14. Panigrahi, B., Shailendra, S., Rath, H.K., Simha, A.: Universal caching model and markov-based cache analysis for information centric networks. *Photon Netw. Commun.* **30**(3), 428–438 (2015). <https://doi.org/10.1007/s11107-015-0570-7>
15. Shailendra, S., Sengottuvelan, S., Rath, H.K., Panigrahi, B., Simha, A.: Performance evaluation of caching policies in NDN—an ICN architecture. In: *Proceedings of the 2016 IEEE Region 10 Conference (TENCON)*, Singapore, 22–25 November 2016, pp. 1117–1121 (2016)
16. Aubry, E., Silverston, T., Chrisment, I.: SRSC: SDN-based routing scheme for CCN. In: *IEEE Conference on Network Softwarization (NetSoft)*, pp. 1–5 (2015)
17. Torres, J., Ferraz, J., Duarte, O.: CROS-NDN: controller-based routing strategy for named data networking. *Electr. Eng. Program.* 1–5 (2014)
18. Aubry, E., Silverston, T., Chrisment, I.: Implementation and evaluation of a controller-based forwarding scheme for NDN. In: *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, pp. 144–151 (2017). <https://doi.org/10.1109/AINA.2017.83>
19. Nguyen, X.N., Saucez, D., Turletti, T.: Efficient caching in content-centric networks using OpenFlow. In: *Proceedings - IEEE INFOCOM*, pp. 1–2 (2013)
20. Van Adrichem, N.L.M., Kuipers, F.A.: NDNFlow: software-defined named data networking. In: *1st IEEE Conference on Network Softwarization: Software-Defined Infrastructures for Networks, Clouds, IoT and Services, NETSOFT 2015* (2015)

21. Mccauley, J.: Pox: a python-based openflow controller (2014). <https://github.com/noxrepo/pox>
22. Riley, G.F., Henderson, T.R.: The ns-3 network simulator. In: Wehrle, K., Güneş, M., Gross, J. (eds.) *Modeling and Tools for Network Simulation*, pp. 15–34. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12331-3_2
23. Putra, M.A.P., Situmorang, H., Syambas, N.R.: Least recently frequently used replacement policy named data networking approach. In: *Proceedings of the 2019 International Conference on Electrical Engineering and Informatics (ICEEI)*, Bandung, Indonesia, 9–10 June 2019, pp. 423–427 (2019)
24. Tody Ariefianto, W., Syambas, N.: Routing in NDN Network: A Survey and Future Perspectives (2017). <https://www.semanticscholar.org/paper/Routing-in-NDN-network>. Accessed 20 Jan 2022
25. Zhang, Q.-Y., Wang, X.-W., Huang, M., Li, K.-Q., Das, S.K.: Software defined networking meets information centric networking: a survey. *IEEE Access* **6**, 39547–39563 (2018). <https://doi.org/10.1109/ACCESS.2018.2855135>



New Approach to Telecom Churn Prediction Based on Transformers

Jalal Rabbah^(✉) , Mohammed Ridouani^{}, and Larbi Hassouni^{}

RITM Laboratory, CED Engineering Sciences, Hassan II University, Casablanca, Morocco
r.jalal@gmail.com

Abstract. Retaining customers costs more than acquiring new ones to telecommunications providers. Predicting churn has therefore become an extremely critical issue for leading service providers worldwide. As a result, major investments are increasingly being made in the design of innovative strategies to combat churn. Machine learning is part of the efforts performed nowadays in this field. In the present paper, we have built a new approach to predict with a high accuracy the churn, our model is based on transforming features into Radar Chart image and then apply a transformer architecture for classification. The model we suggested scored a good result in terms of accuracy, 81%.

Keywords: Transformers · Deep learning · Radar chart · Customer churn · ViT

1 Introduction

In the era of Big Data, many companies are setting up their own data collection and processing systems to understand their customers' behavior and thus opt for new “data centric” marketing strategies. However, while many marketers are happy to observe their customers' behavior, some say they don't have the keys to act on the churn phenomenon. So what is churn and how can new artificial intelligence technologies help anticipate and reduce it? This is the subject we are going to discuss in this paper.

The term “churn”, widely used in marketing, refers to a downward trend in customer numbers. This concept, also known as “attrition”, allows us to analyze customer loyalty and the impact of marketing actions on them. The churn rate is calculated using the following simple formula: it consists of the division of the count of lost customers by the total customer base.

Churn rate has a direct impact on customer lifetime value (CLTV). That's why focusing efforts on reducing churn will not only improve CLTV but also generate a better gain regarding all the costs saved for customer acquiring. Then the reduction of churn is among the best ways to boost subscription system performance for any kind of business. Among the sectors most affected by the churn phenomenon is telecommunications industry. For which, retaining customers is important for saving revenues. In order to address this situation, telecom service companies are challenged to develop retention programs and to focus more on customer retention than on new customer acquisition.

As opposed to large multiple Terabytes of data, typically holding multimedia data feeds, telco data stores commonly contain numerous records of network data records. Thus, the task of handling this enormous amount of data is a large one that needs specific knowledge and implication. Among the greatest obstacles for solving such problem is the required time for computing to detect genuine churn; In the telecommunications industry, it's not unusual for companies to see churn rates as high as 7–10%. For this reason, churn prediction is an important business practice. Churn prediction helps telecom companies identify which customers are most at risk of leaving and why. When companies can anticipate which customers are leaving, they can take steps to reduce churn and increase profitability. If the churn prediction model is accurate, it can help telecom companies be proactive and stop churn before it happens. They can use this model to send offers or messages to customers who are at risk of leaving. This can help retain current customers and increase revenue.

Forecasting churn is predominantly a classification task, where the objective is to anticipate pre-departure churn. Specifically, we consider binary labels such as Target for each subscriber. Note that we don't have the same percentage of users, only few customers may be churners in a given sample of data. This type of figures is a typical unbalanced data sets use case. For process purpose and to have acceptable results, it is important to consider several characteristics to make enough visibility on all eventual factors leading the customer to quiet.

Many approaches have been implemented to choose the most efficient way to predict churn, Sudharsan, R., & Ganesh, E. N [1] proposed a Swish RNN based customer churn prediction, De, S., Prabu, P., & Paulose, J [2] discussed various approaches to churn prediction, including machine learning, data mining, and customer segmentation. Lalwani, P., Mishra, M [3], presented a new machine learning approach to detect churners using ensemble learning. Pejić Bach, M., Pivar, J. [4] presented an Approach that merged decision trees and cluster analysis. In previous papers [5] we suggested a prediction model based on multi-layer Stacking architecture and recently, we published a new architecture based on features transformation to images, combined with convolutional neural networks architecture and stacknet. H.Abbasimehr, [7], evaluated well known techniques of ensemble learning (Stacking, Bagging, Voting and Boosting).

In summary, this research objective is to predict churn by building a supervised model able to identify quickly and accurately churners. This work is as composed of six sections: Sect. 2 for problem description. Section 3 shows the vision transformer used in our study. Section 4 describes the model architecture. Section 5 gives the results of the experiments. And we concluded with Sect. 6.

2 Description of the Problem

There are several reasons why customers may churn, such as dissatisfaction with the service, poor customer service, or the availability of a better offer from a competitor. Churn can be a particularly challenging problem to address because it is often difficult to predict which customers are at risk of churning, and it can be costly to try to retain customers who already plan to leave (Fig. 1).

To address this problem, telecommunications companies often use data analysis and machine learning techniques to identify patterns and trends that may indicate a

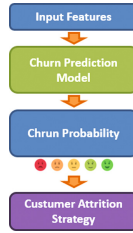


Fig. 1. Customer attrition process

customer is at risk of churning. This allows the company to proactively address the issue and potentially retain the customer. In addition, companies may use customer surveys, focus groups, and other methods to gather feedback and identify the root causes of churn. By understanding the reasons why customers are leaving, companies can take steps to improve their products, services, and customer experience to reduce churn and improve retention.

Significant Features, Behaviors and Data Volume, Time and Money are the main challenges that telco companies and practitioners are facing to find a definitive and radical solution to this problem. Which remains despite the multiple works already published a very promising field of research.

3 Vision Transformers

Vision transformers [8] models split images into a series of positional embedding patches that the transformer encoder will process one by one. This architecture has been proposed primarily for block-to-block learning [9], as well as for automatic translators. With great efficiency, transformers then became the preferred choice for natural language processing applications. [10], Nevertheless, the dominant computer vision architecture has remained based on CNNs [11, 12, 23], many scholars proposed to substitute the convolution with the self-attention [13]. Yet, the use of specific designs in the attention makes it difficult to scale the models on hardware acceleration devices [14]. Theoretically, self-attention is capable of being trained to behave in a similar way to convolution. Empirically, the small sized input image patches, make this model relevant only to low resolution images. With no particular restrictions on the size of patches, using ViTs (vision transformers) it is possible to retrieve patches from images and pass them into a transformer encoder to get a comprehensive representation, ultimately to be transformed for classification [15]. Remarkably, transformers demonstrate improved scalability compared to CNNs: Vision Transformers significantly surpass the ResNets model when training larger models on larger datasets. Like a network architecture design landscape in natural language processing, transformers have also changed the game in computer vision. For this reason, we have transformed churn dataset features to images, to take advantage of the strength of this type of models which is able to distinguish behaviors with a very acceptable precision (Fig. 2).

In this paper, we have used the hugging face transformers library [16] for pytorch [17]. This library gives access through APIs and tools to easily download and train

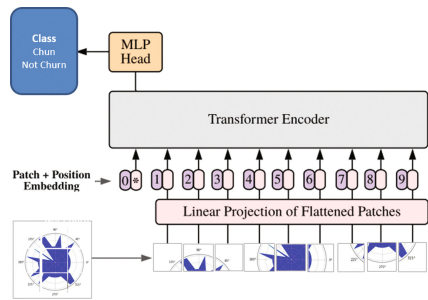


Fig. 2. Model architecture

advanced pre-trained models. Working with pre-trained models saves computational costs, reduces the carbon footprint, and the time and effort needed to train a model from the ground up. These models support standard tasks in different modalities. We have used the Computer Vision and image classification model called ViTModel.

4 Methodology

4.1 Step 1: Data Understanding

To build our model, we used a customer churn dataset from telecommunications sector, containing 7043 customers defined with 21 dimensions or features shown in the table below (Table 1) and downloaded from Kaggle [18] datasets.

4.2 Step 2: Data Preparation

Various manipulations and transformations have been made to our raw dataset, in order to build our model, data was first normalized and then encoded features using zeros and ones for boolean features and specific codes for non-Boolean. A Summary descriptive statistic representation of our dataset is shown in Table 1.

Table 1. Used dataset to build the churn prediction model

1-features	#	Mean	Std	Min	25%	50%	75%	Max
Gender	7032	0.505	0.5	0	0	1	1	1
SeniorCitizen	7032	0.162	0.369	0	0	0	0	1
Partner	7032	0.483	0.5	0	0	0	1	1
Dependents	7032	0.298	0.458	0	0	0	1	1

(continued)

Table 1. (continued)

l-features	#	Mean	Std	Min	25%	50%	75%	Max
Tenure	7032	32.422	24.545	1	9	29	55	72
PhoneService	7032	0.903	0.296	0	1	1	1	1
OnlineSecurity	7032	0.287	0.452	0	0	0	1	1
OnlineBackup	7032	0.345	0.475	0	0	0	1	1
DeviceProtection	7032	0.344	0.475	0	0	0	1	1
TechSupport	7032	0.290	0.454	0	0	0	1	1
StreamingTV	7032	0.384	0.486	0	0	0	1	1
StreamingMovies	7032	0.388	0.487	0	0	0	1	1
PaperlessBilling	7032	0.593	0.491	0	0	1	1	1
MonthlyCharges	7032	64798	30.086	18.25	35.588	70.350	89.862	118.75
TotalCharges	7032	2283.3	2266.771	18.80	401.450	1397.475	3794.7	8684.8
Churn	7032	0.266	0.422	0	0	0	1	1

4.3 Step 3: Feature Engineering

During feature engineering, we have performed a feature selection study to define most significant features to reduce dimensionality which is important for performance and results quality. To achieve this task, we have made a scoring of our features using different algorithms, this permits to us to have the results presented in the below table (Table 2).

Table 2. Features selection study output

	1_RFE	2_Chi-2	3_Pearson	4_Logistics	5_RandomF	6_LightGBM
1_tenure						
2_TotalCharges						
3_Tenure_0-12_TenureG						
4_Monthly_Charges						
5_InetServ_Fiber_Optic						

Using 6 methods for feature selection, we were able to select only most important 15 features and delete the less significant features based a predefined threshold. For each algorithm we obtained a score for each feature, we then calculated for each feature the average of the scores resulting from the 6 algorithms. Finally we applied a threshold to choose only the features verifying the threshold.

Below is the list of used algorithms:

- Pearson correlation
- Chi-Squared
- RFE
- LASSO
- Random Forest
- Logistic Regression

In summary, after the previous steps results a new dataset with a reduced dimensionality and only containing relevant features, less important information has been eliminated to make our model more performant. Below graphs (Fig. 3 & Fig. 4) can clearly explain the behaviors of churn and non-churn subscribers. And explain then why some features are not relevant for our use case. For instance, some features have no significant meaning to our problem such as Multilines_NoPhoneSers, Multi-lines_Yes and DeviceProtection.

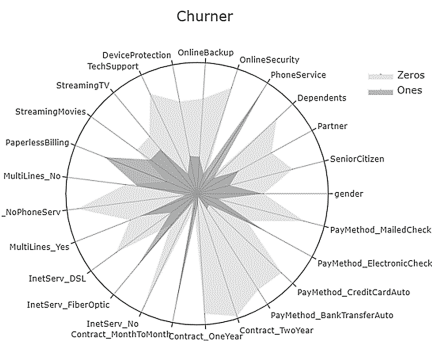


Fig. 3. Churners feature distribution

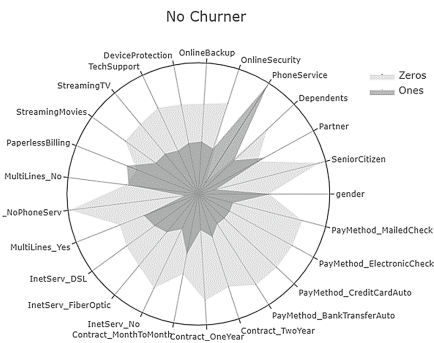


Fig. 4. No churners feature distribution

4.4 Step 4: Features Transformation to Images - Radar Chart

Radar plotting refers to a visual representation of multi-variate space information as a 2-dimensional plot of at least three variables on axes originating at a single point in the data. While the location and relative angle of the axis are generally uninformative, various heuristics, like algorithms that plot the data as a total area, may be used to order the variables for their relative locations that show correlations, biases, and a variety of additional comparison metrics.

To create images (Fig. 5) we have used the python package called plotly [19] to transform the prepared churn dataset to images and matplotlib [20] to export images.

4.5 Step 5: Model Building

The following section will present the architecture of our model called ChurnViT.

This type of use case frequently faces such a challenge of unbalanced dataset. Telecom churn prediction is one of the cases where positive cases are very low compared to most subscribers.

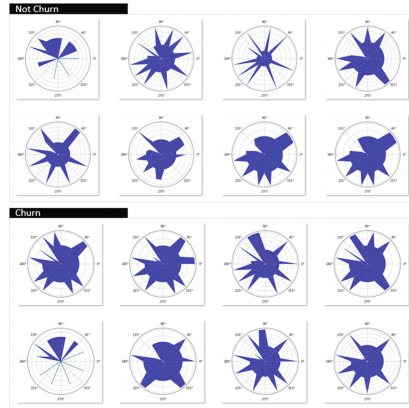


Fig. 5. Radar Chart generated image from churn dataset features transformation (images size: 224×224)

One of the most well-known methods for dealing with unbalanced datasets is the Synthetic Minority Oversampling Technique (SMOTE) [21].

After Data preparation and enrichment, we transformed the churn dataset features to radar charts and then exported as images. As already presented, we have used a pre-trained model [14] from Huggingface opensource initiative to create a churn prediction model based on the generated images. This model was pre-trained on 14 million 224×224 images and more than 21 thousand classes on ImageNet-21k. For training and classification, Images are presented to the model as a sequence of fixed-size 16×16 resolution patches. A high-level overview of our churner's prediction architecture is presented in the Fig. 6.

The first step is to clean the raw data and then verify missing values and fix all data quality issues. The second step is to check and remove outliers. After data cleaning step, comes the data preparation and the optimization phase, such as normalization, rescaling, and transforming categorical to numerical.

The third step is feature selection, as mentioned below we have combined many algorithms and technics to score the features and finally choose the ones with the highest scores. The remaining least relevant features are deleted. One we applied the SWOT technique to balance the dataset the next step is to transform the generated data to radar charts and exported as images [22]. The set of images was then used for training of our Vision Transformer (EPOCHS = 3 and LEARNING RATE = 2.10^{-5}), we also used the remaining data to validate and test our architecture.

5 Results of Experiments

5.1 Results

To assess the performance of our design, we used the following metrics:

- Classification accuracy or the proportion of correct predictions out of all predictions

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (1)$$

- Precision or the proportion of accurate positive predictions

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

- Recall or hits rate

$$Recall = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negative}} \quad (3)$$

- F-measure:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

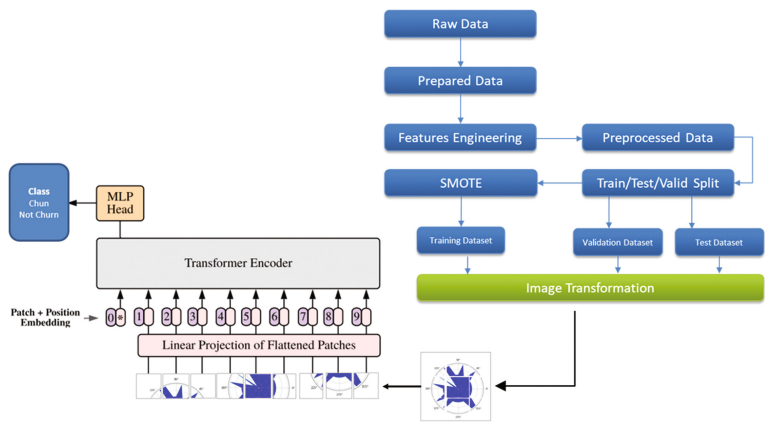


Fig. 6. The churn prediction system based on image transformation and ViT

Table 3. Represent the obtained performance of our model and a comparison to other sta-dies:

	ChurnViT	Stacknet [6]	Decision tree	Logistic regression	Random forest
Accuracy	81%	83%	73%	79%	79%
Recall	95%	82%	80%	89%	79%
F1-score	87%	81%	83%	86%	77%

In Table 3, we presented the results obtained using our model named ChurnViT, tested using the dataset presented in Sect. 4.1. We also put in the same table results using other models. Simple architectures (mono model) such as decision tree, Random fores give less performance then more complex architectures like ensembles or transformers.

5.2 Discussion

In this paper we used a clean dataset from the internet containing 7043 data points with 27% churners, we obtained 81% accuracy to predict churners using this dataset. This result is quite excellent in the context of marketing and customer management, for which such score can avoid important amount of money due to churn. Whereas the recall and F1-score of our model are the best among the other models of the list, this means that we have a very low false negative rate and very low false positive rate, that means we can correctly identify real threats and our model is not disturbed by false alarms. We can improve the performance by adding other features such as customer care interaction information or by using more historical data regarding customers and Business ecosystem in general. The next step is to find out when and why churn occurring according for example to a features importance analysis.

The results show that the approach using transformers is more accurate and more suitable to our use case, it can even be improved looking at the potential that vision transformers offer in the field of image classification.

6 Conclusion

This work evaluated a new way of predicting churn using digital image transformation against different types of classification algorithms. According to the obtained results, which can be further improved by using other ViT architectures, the suggested classifier can provide a suitable response for telecommunication service providers in order to get early and accurate alerts about potential churners, which will allow them to launch focused retention and loyalty campaigns, as well as saving a significant budget that would eventually be wasted on both customer departure and new customer acquisition expenses.

References

1. Sudharsan, R., Ganesh, E.N.: A swish RNN based customer churn prediction for the telecom industry with a novel feature selection strategy. *Connect. Sci.* **34**(1), 1855–1876 (2022)
2. De, S., Prabu, P., Paulose, J.: Effective ml techniques to predict customer churn. In: 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 895–902. IEEE, September 2021
3. Lalwani, P., Mishra, M.K., Chadha, J.S., Sethi, P.: Customer churn prediction system: a machine learning approach. *Computing* **104**(2), 271–294 (2022)
4. Pejić Bach, M., Pivar, J., Jaković, B.: Churn management in telecommunications: hybrid approach using cluster analysis and decision trees. *J. Risk Finan. Manag.* **14**(11), 544 (2021)
5. Rabbah, J., Ridouan, M., Hassouni, L.: A new telecom churn prediction model based on multi-layer stacking architecture. In: Ben Ahmed, M., Abdelhakim, B.A., Ane, B.K., Rosiyadi, D. (eds.) *Emerging Trends in Intelligent Systems & Network Security. NISS 2022. Lecture Notes on Data Engineering and Communications Technologies*, vol. 147, pp. 35–44. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-15191-0_4
6. Rabbah, J., Ridouani, M., Hassouni, L.: A new churn prediction model based on deep insight features transformation for convolution neural network architecture and stacknet. *Int. J. Web-Based Learn. Teach. Technol. (IJWLTT)* **17**(1), 1–18 (2022)

7. Abbasimehr, H., Setak, M., Tarokh, M.J.: A comparative assessment of the performance of ensemble learning in customer churn prediction. *Int. Arab J. Inf. Technol.* **11**(6), 599–606 (2014)
8. Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S., Shah, M.: Transformers in vision: a survey. *ACM Comput. Surv. (CSUR)* **54**(10s), 1–41 (2022)
9. Brown, T., et al.: Language models are few-shot learners. In: *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901 (2020)s
10. Devlin, J., Chang, M. W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)* (2018)
11. Techa, C., Ridouani, M., Hassouni, L., Anoun, H.: Alzheimer's disease multi-class classification model based on CNN and StackNet using brain MRI data. In: Hassanien, A.E., Snášel, V., Tang, M., Sung, T.W., Chang, K.C. (eds.) *Proceedings of the 8th International Conference on Advanced Intelligent Systems and Informatics 2022. AISI 2022. Lecture Notes on Data Engineering and Communications Technologies*, vol. 152, pp. 248–259. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-20601-6_23
12. Elaanba, A., Ridouani, M., Hassouni, L.: Automatic detection using deep convolutional neural networks for 11 abnormal positioning of tubes and catheters in chest X-ray Images. *IEEE World AI IoT Congress (AIIoT) 2021*, 0007–0012 (2021)
13. Ramachandran, P., Parmar, N., Vaswani, A., Bello, I., Levskaya, A., Shlens, J.: Stand-alone self-attention in vision models. In: *Advances in Neural Information Processing Systems*, vol. 32 (2019)
14. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint [arXiv:2010.11929](https://arxiv.org/abs/2010.11929)* (2020)
15. Cordonnier, J.B., Mahendran, A., Dosovitskiy, A., Weissenborn, D., Uszkoreit, J., & Unterthiner, T.: Differentiable patch selection for image recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2351–2360 (2021)
16. Wolf, T., et al.: HuggingFace's transformers: state-of-the-art natural language processing. *arXiv preprint [arXiv:1910.03771](https://arxiv.org/abs/1910.03771)* (2019)
17. Chintala, S.: An overview of deep learning frameworks and an introduction to pytorch (2017)
18. BlastChar, Telco Customer Churn Dataset (2018). <https://www.kaggle.com/blastchar/telco-customer-churn>. Accessed 23 Feb 2018
19. Rhett, A.: Mass of 8 x 11.5 Sheets of Paper. Plotly Technologies Inc., <https://plot.ly/~RhettAlain/412/mass-of-8-x-115-sheets-of-paper/>. Accessed 21 Apr 2015
20. Hunter, J.D.: Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**(3), 90–95 (2007)
21. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002)
22. Benazzouza, S., Ridouani, M., Salahdine, F., Hayar, A.: A novel prediction model for malicious users detection and spectrum sensing based on stacking and deep learning. *Sensors* **22**(17), 6477 (2022)
23. Rabbah, J., Ridouani, M., Hassouni, L.: A new classification model based on stacknet and deep learning for fast detection of COVID 19 through X rays images. In: *2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)*, pp. 1–8. IEEE, October 2020



Smart Healthcare Development Based on IoMT and Edge-Cloud Computing: A Systematic Survey

Fatima Ezzahra Moujahid^(✉), Siham Aouad, and Mostapha Zbakh

National School of Computer Science and Systems Analysis, Mohamed V University
in Rabat, Rabat, Morocco

fatimaezzahra_moujahid@um5.ac.ma,
{siham.aouad,mostapha.zbakh}@ensias.um5.ac.ma

Abstract. Smart healthcare has become one of the important themes of our time. The growth and heterogeneity of medical data are seen as major challenges, especially in terms of processing, performance, storage, and security, which has made it obvious and essential to use various technologies to make smart healthcare more efficient and convenient, including the Internet of Medical Things (IoMT), big data, and cloud computing. However, these technologies may pose new challenges in terms of bandwidth cost due to the number of transfers to be made between a number of nodes, security of a large amount of medical data, and delay, especially since time is a very important factor in saving patients' lives. Some papers propose the joint use of these technologies to mitigate these challenges. In this paper, we introduce the concept of intelligent healthcare, and then we present the recent solutions to address the mentioned challenges. Finally, we draw a conclusion and some perspectives for this work.

Keywords: Internet of Medical Things · Edge-cloud computing · Artificial Intelligence · Smart healthcare · Natural language processing

1 Introduction

Smart Healthcare was born in 2009 from the Smart Planet concept at IBM (Armonk, NY, USA) [1]. A Smart Planet is an intelligence infrastructure that uses miscellaneous sensors to collect information, transmit it via the Internet of Things (IoT) and process it using cloud computing. The use of new technologies such as Internet of Medical Things (IoMT) is having a significant impact, building on the platform of wearable devices, these IoMT devices are equipped with sensors that generate large amounts of data that are transmitted via the internet infrastructure to a cloud computing, where the dense processing of the data is performed [2].

Smart healthcare has shifted from disease prevention to disease treatment, it helps doctors to prevent life-threatening diseases even before they occur, it also provides communication technology to interact between medical institutions, doctors, nurses, patients and multiple medical devices [1].

Artificial Intelligence (AI) can simplify lives of patients, doctors, nurses and medical staff by accomplishing tasks normally done by humans, at lower cost and in less time. It also helps them respond faster with AI solutions for healthcare. AI can identify problems and quickly alert care teams, allowing providers to discuss options and make vital treatment decisions faster.

Given the magnitude of medical data, IoMT technologies alone cannot meet all the needs, especially in terms of real-time processing, analysis and response, which can put patients' lives at risk. hence the use of cloud computing become necessary, cloud can provide powerful computing resources with huge storage capacities which helps to solve resource level problems, however the cloud is characterized by high transmission costs, limited coverage and long response time [3]. IoMT-based solutions provide real-time response by solving the centralization problem that exists in the cloud, enabling the reduction in latency required for time-dependent solutions such as vital signs monitoring [4,5].

Due to the use of cloud and edge computing, patient information is vulnerable to various attacks such as denial of service, data corruption, privacy disclosure, malicious gateways, etc. In order to reduce the risks, it is important to use security protocols and algorithms such as hash function, encryption algorithm and attribute encryption algorithm, as well as develop isolation mechanisms for different edge nodes.

Cloud computing is the use of centralized and remote servers over the Internet to store, process, and analyze data. However, edge computing adopts a decentralized hierarchical design, which can use discrete computers and devices as edge servers [6]. Because of this difference in architecture, edge computing can provide bandwidth, location, real-time, low-cost and privacy services compared to cloud computing. Therefore, although edge and cloud computing are the same in terms of functions they provide, the means of attack are very different [7]. For example, the effort of Distributed Denial of Service attack (DDOS) is more important in a cloud computing system because of the difference in the protocols used. Indeed the cloud uses strong security protocols like HTTP/HTTPS, while the edge adopts light network protocols like UDP, however, The edge is vulnerable to exploiting vulnerabilities in wireless authentication protocols such as Bluetooth and 4G/5G, while the cloud is vulnerable to exploiting vulnerabilities in protocols such as SSL/TLS that can be applied to wired protocols [7].

In this paper, we examine various related research areas. IoT, IoT data fusion, cloud, edge and AI in smart healthcare, privacy and issues in IoT-based smart healthcare. Section 2 provides a comparison between existing research and focuses on IoMT and AI in healthcare. In Sect. 3 we discuss all of the previously cited works according to the presented metrics. Section 4 presents some challenges and future directions. Finally, we draw a conclusion.

2 Literature Review

The IoMT is a combination of medical devices and the IoT. It is a healthcare system, where every medical device is monitored and connected over the Internet [8]. However, it is hard to establish connections between multiple medical institutions, Due to the high demands on data storage, management and processing capabilities, as well as the difficulty for IoMT to guarantee integrity and confidentiality of medical data. Then the use of cloud computing become a necessity due to its high sharing, reliability, and scalability, even though the use of cloud alone results in significant delay and higher energy consumption due to the high load. Which leads to think of combining cloud and edge computing in order to obtain real-time results especially when this is related to the life of the patients [9].

2.1 Architecture, Technology and Application

IoMT architecture is composed of three layers namely [9]: The perception, the network, and the application layer. The first one can be divided into two sub-layers: the data acquisition sub-layer, which uses signal acquisition methods such as Radio Frequency Identification (RFID) technology, and the data access sub-layer, which connects the data between the data acquisition and the network layer. The second one is composed of: the network transmission layer, which transmits the data collected by the perception layer using the Internet and other special networks in real-time, and the service layer, which enables the integration of heterogeneous networks, different data formats and warehouses. The third one is made up of: the medical information application, such as patient information and medical equipment management, and the medical information decision-making application, which includes disease information analysis, diagnosis and treatment. IoMT technology is composed of three technologies namely [9]: RFID, this technology uses radio frequency to read and write data, as well as it guarantees bidirectional data communication without contact and does not require human intervention, RFID is generally used in medical asset management and vital signs data collection. The Wireless Sensor Network (WSN) represents a wireless network system, this technology allows to monitor, perceive and collect information in real-time and send it wirelessly. WSN represents advantages namely speed, reliability, security, as well as real-time response, and middleware, which is considered to be software that serves as a gateway between applications, it allows to support distributed computing and cope with detection devices, realizes the standardization of various application as well as communication of data between application systems. IoMT is often used in healthcare, it enables timely understanding of the patient's physical condition [10] based on real-time vital signs monitoring [11–14]. It allows analysis of data, diagnosis of diseases, and decision making. The IoMT also has services to improve, namely the flexibility, design speed and interconnection of portable devices [9].

Another newly emerging cloud-based IoMT application area is called Cloud-Enabled IoMT, its architecture is divided into three layers [9]: Medical service, Service management, and User layer. The medical service layer contains Medical infrastructure and provides high-performance computing, network and storage resources and services based on virtualization, Medical Platform provides a platform or technical support that enables users to develop their own applications, and Medical software ensure the security and confidentiality of patient data, as well as provides demands to both the service management layer and the platform as a service layer for development. The Service management layer provides technologies and functions and ensures the medical system's safety and stability. User layer allows users to connect to the cloud and use its resources, as well as providing them with the possibility of transmitting their own resources and data in order to guarantee the sharing of information with doctors, thus eliminating risks to patients' lives. Cloud-based IoMT technologies are divided into [9]: Cloud computing technology is based on the principle of virtualization, the cloud guarantees infinitely scalable resources in computation, storage, and network, it enables the pooling of service resources, performance, resource sharing and high reliability, but the cloud has security issues, which exposes patient data to threats. Then Big Data technology that can control the huge amount of data, process the data professionally, collect and preprocess the data, store it, analyze it and exploit it. The data is divided into 3 types structured using MySQL, Oracle for storage, semi-structured using Hadoop and unstructured using NoSQL. Lastly Artificial Intelligence (AI): the most used categories in AI are: Machine Learning (ML) which creates mathematical models from data. This technology helps the medical profession to predict and diagnose diseases and make decisions, Deep Learning (DL) which is mainly based on neural networks, helps to improve the efficiency of diagnosis and treatment of diseases, Natural Language Processing (NLP) which converts unstructured data text into structured data text, as well as the use of robots facilitates the tasks for doctors and nurses. Cloud-based IoMT reduces the cost of data processing, improves medical services efficiency, and solves the problem of islands of information in medical facilities by using an electronic medical record-based data-sharing platform. The cloud suffers from security issues, which requires the use of other applications and algorithms, for example, IBBE algorithm that encrypts patient data before storing it in the cloud, and physicians have to decrypt it to obtain the data. The proposed LPP-MSA scheme in [15] adopts an online/offline signature and server-based verification mechanism to ensure patient privacy with low storage and computation costs.

The Edge-Cloud IoMT is considered as another area that combines the use of the cloud and the edge, its architecture is divided into three layers [9]: The terminal, the edge computing and The cloud computing layer. The first layer which consists of devices such as wearable devices and sensors, this layer collects data from the local device and uploads it to the edge device. The second layer which consists of nodes such as smartphones, tablets, gateways or routers, which provides edge computing, network and storage services for the data. This layer

reduces the delay as well as security risks on the data. The third layer can be considered as a giant processing center, allowing processing, storing and analyzing data downloaded by the edge computing layer. Edge-Cloud IoMT technologies are categorized as follows [9]: Edge computing a distributed information technology architecture where data is analyzed and processed at the edge of the network, close to the source of the data with the essential network, computation, and storage capabilities. This technology connects the end devices with the cloud which guarantees the passage of data in a bidirectional manner. Edge computing has advantages such as programmability, high definability, security, and low delay. Edge computing is composed of 3 categories, mobile edge computing (MEC) [16,17] which exploits the mobile network edge to enable cloud computing capabilities to nearby users, cloud computing (MCC) which can be considered as a small mobile data center, and fog computing [18,19] consists of low-performance and decentralized computing devices [9]. Edge-Cloud IoMT helps medical personnel improve the efficiency and accuracy of healthcare, providing the right information at the right time, which can be life or death related. This technology can process, handle, and store patient data collected locally, even offline, facilitating remote patient monitoring [20]. Using edge computing, AI, and other technologies, a health monitoring architecture can be used to collect data on various vital signs to aid in diagnosis [21]. With edge computing, the healthcare system can optimize data transmission and reduce delays and energy consumption [22]. Edge-Cloud IoMT facilitates interaction between patients and their doctors, making Edge-Cloud IoMT a decision support tool [9].

To expand the traditional IoMT and maximize its efficiency, the authors of [9] proposed to introduce cloud computing and edge computing at the architecture, technology, and application levels. Table 1 shows a comparison between traditional IoMT, Cloud-Based IoMT, and Edge-Cloud IoMT.

2.2 AI in Healthcare

AI refers to a computerized system capable of performing cognitive functions and physical tasks, solving different problems, or making decisions without explicit human instruction [23], using several methods in healthcare applications and services such as [24] Machine Learning (ML) which introduces a field of AI that can be divided into 3 approaches: Supervised, Unsupervised and Reinforcement learning algorithms. The first two are based on a mathematical model, and the main difference between the two methods is the use of labeled data sets. Supervised learning uses labeled input and output data [25], while unsupervised learning algorithms do not use labeled output data [26]. Reinforcement learning algorithms do not require knowledge of mathematical model of the discrete-time stochastic control process. The Neural networks (NNs) is also used in the medical field, it consists of digitized inputs that pass through several hidden layers of connected artificial neurons, with each layer responding to a different feature to incrementally identify the features and provide an output. Deep neural networks (DNNs) are often used when data must be interpreted to infer specific patterns from different types of clinical images and to discover patterns from data sets

Table 1. Comparaision between IoMT, Cloud-Enabled IoMT and Edge-Cloud IoMT

	IoMT	Cloud-Enabled IoMT	Edge-Cloud IoMT
Architecture	-Perception layer (Data acquisition sublayer-Data access sublayer) -Network layer (Network transmission layer-Service layer) -Application layer (Medical information application-Medical information decision-making application)	-Medical Service Layer (Medical infrastructure-Medical platform-Medical software) -Service Management Layer -User Layer	-Terminal layer -Edge computing layer -Cloud computing layer
Technology	-Radio Frequency Identification (RFID) -Wireless Sensor Network (WSN) -Middleware	-Cloud Computing -Big Data -Artificielle Intelligence (AI)	-Edge Computing (Moving Edge Computing (MEC)-Micro Cloud Computing (MCC)-Fog computing) -Computation Offloading
Application	-Timely understanding of the patient's physical condition -Analyze and make decisions on information -Optimize the hospital business process -Improves efficiency and resource utilization -Problem of the information island in the different medical institutions	-Problem of the information island in the different medical institutions -Patient privacy issues -Problem of managing exponential data growth -Centralized system (send data and wait for results, network failure)	-Improves work efficiency of medical staff Process, calculate and store the data collected by the edge devices in real-time at the local level (decentralized distributed system) -Useful for remote patient monitoring -Reduce energy consumption and optimize data transmission -Reduces the data flow generated by the network -Guarantee the security of stored data

such vital signs [24]. Deep Learning, differs from traditional Machine Learning in the way representations are learned from raw data. Indeed, DL enables computational models composed of multiple processing layers based on Neural Networks to learn data representations with multiple levels of abstraction [27], the main difference between DL and traditional Artificial Neural Networks (ANNs) is the number of hidden layers, their interconnection, and the ability to learn significant abstractions from the input [28]. Natural Language Processing (NLP) is a subfield of AI that can be defined as a set of theoretical computational techniques for analyzing and representing naturally occurring text at one or more levels of linguistic analysis to achieve human-like language processing to accomplish a A series of tasks or applications [29], it enables computers to read and understand human language, and enables natural interactions between computers and humans, enabling computer systems to understand and manipulate natural language to perform needed tasks [30].

AI can be applied to the detection and diagnosis of diseases and medical problems, which is one of the challenges in applying machine learning to healthcare. AI also exists in personalized medicine or more effective treatments based on personal health data and predictive analytics. This range is currently determined by supervised learning, allowing doctors to choose from a more limited set of diagnoses. Machine learning aids in drug discovery, early initial drug screening and preliminary testing can use ML systems, and methods for predicting drug success taking into account various biological factors. Machine learning systems in surgical robots are one way to integrate automation into future healthcare systems. Technologies such as machine learning and deep learning could one day fully automate surgical robots, completely excluding humans from the surgical procedure. Machine learning enables intelligent electronic medical records to facilitate record keeping, explain health conditions and propose treatment plans [31].

2.3 IoMT, Cloud and Edge Computing in Healthcare

IoT is a network of devices with sensors connected to the Internet and allow to exchange data. The Internet of Medical Things has been extensively used by healthcare practitioners and organizations to remotely monitor patients and manage equipment such as Computed Tomography (CT), Positron Emission Tomography (PET), and radiographic (X-ray) machines. The IoMT is also expanding to tracking equipment inventory in individual hospitals, and even medical tools equipped with sensors. Some facilities use sensors to track patients and staff.

Wearables devices connected to smartphones are an important part of IoMT. Healthcare already uses these devices to monitor vital patient activities such as blood pressure and heart rate. However, with the popularity of smartphones and the increased availability of wearables and remote monitoring devices, there is growing consumer interest in these products. Wearable device usage increased from 9% in 2014 to 33% in 2018, while mobile and tablet health applications usage increased from 16% to 48% over the same period [32].

IoMT has many applications in healthcare, this includes improving the patient experience, with seamless connectivity between devices allowing patients to control room temperature and lighting, communicate via video calls, and call nurses via intercom [33]. Healthcare professionals can also track patient movements, and IoT devices also aid in real-time environmental monitoring [34]. IoMT can help in preventing infection, indeed hand hygiene monitoring systems help define and monitor the cleanliness of medical staff. The simplest function of a hand hygiene IoT device is to beep when a medical worker approaches a patient's bedside without washing their hands. IoMT also supports continuous monitoring to provide patients with appropriate healthcare. IoMT devices can apply and analyze complex algorithms. This helps reduce the risk to the patient's life [35].

The study [36] used a multisensor platform with a dual-channel pulse wave (PPW) signal and a single-channel Electrocardiogram (ECG) to estimate blood pressure (BP). After collecting the signals, the features containing the needed parameters are extracted, and then the most important ones are selected. The features are fused and the blood pressure models are improved using a multi-instance regression algorithm, the estimates proposed by the models based on this algorithm and the cuff-based values of 85 participants allow concluding that the estimated error is equal to 1.5. The complexity of this method depends on the number of channels used, the frequency transformation for each signal, the number of data points denoted by T in the training dataset, and the number of features extracted for selection. And since all parameters are negligible in front of T , then the complexity is $O(T^2)$.

In [37], the authors proposed a technique to recognize and classify emotions. Using two public datasets, DEAP and SEED, the study involved recording participants' physiological signals while watching music videos, it consists of forming a vector containing the extracted features fused and selected. They improved accuracy by using a combination of ST-SBSSVM significance testing as well as sequential backward selection with SVM. The input methods used include a 32-channel Electroencephalogram (EEG) signal, a four-channel Electrooculogram (EOG) signal, a four-channel Electromyogram (EMG) signal, and vital signal measurement. This method achieves 72% accuracy on the DEAP database and 89% accuracy on the SEED database.

The authors of [38] proposed a real-time monitoring system based on a Regression Forecast-Support Vector Machine (RF-SVM) model used to identify the situation's level and fuse the data to ensure accuracy and reduce the risk to mine workers. They focused on open-pit mines, considering two important indices: slope extent and slope displacement, they used sensors to predict mine stability. Heterogeneous data are preprocessed to reduce noise, fill in missing data using regression algorithms, normalize data, and use information entropy theory to measure uncertainty. The analysis appear a root mean square error (RMSE) of less than 0.2.

The authors of [39] proposed an ensemble approach that merges data collected by a body sensor network (BSN) and wearable sensors, merged and connected to an ensemble classifier using the random forest and kernel random forest algorithm to predict heart disease. The ensembles are placed in a fog computing environment for low complexity processing, then sent to the cloud for computation, analysis and storage using its processing capabilities. The data is pre-processed to reduce noise, to extract the necessary features, to normalize them and select them to avoid overfitting. The predictive accuracy of this method is 98%.

Based on a sensor fusion method, and to ameliorate the reliability and performance of sleep apnea prediction, researchers of [40] have provided a model to collect and integrate data from different sensors using a public dataset SleepHeart-Health-Study-1 database. These data are divided into several categories based on their age, pre-processed, normalized, and the features in question are extracted,

They proposed convolutional neural network (CNN) and long short-term memory (LSTM) models that are considered as basic deep learning (DL) models to reduce noise and analyze the performance of this method using area under the precision-recall (AUPR) with a value of 0.67 for CNN and 0.71 for LSTM.

To ameliorate the detection of mental stress and using a Support Vector Machine (SVM), the authors of [41] fused seven channels of functional near-infrared spectroscopy (fNIRS) with seven EEG electrodes based on a set of 12 individuals with no history of neurological, psychiatric, or drug use illness. Measurements of fNIRS and EEG signals were performed on subjects while they solved arithmetic problems (Montreal Imaging Stress Task) in two conditions of control and stress. Features were fused to construct a vector from the two modalities, and fNIRS and EEG tests were performed in two phases, separately and after fusion. The fusion of fNIRS and EEG signals outperformed each separately.

The authors of [42] proposed an approach using physical concept of potential field, they considered a Magnetic Resonance Imaging (MRI) image as a finite set of pixels, each of them is related to a gray intensity value. They proposed a Potential Field Segmentation (PFS) algorithm for brain tumor segmentation in MRI scans, they used three segmentation methods namely Potential Field Segmentation (PFS), FOR, Potential Field Clustering (PFC), and constructed sets containing MRI images using these segmentations. Two fusion rules are applied: intersection which considers a pixel as a tumor if it is tagged as a tumor in all sets, and union which considers a pixel as a tumor if it is tagged as a tumor only for one of the sets. PFS requires fewer parameters, and resides firstly in calculating the potential field at each pixel, secondly in calculating an adaptive potential threshold, thirdly, each pixel is associated with the tumor set. The study is based on a public database of brain tumor image segmentation (BRATS) and combines the results to obtain a merged segmentation.

In [43] the authors suggested a particle filtering-based approach to ameliorate heart rate monitoring by exploiting existing artifacts and using wearable sensors. This approach fuses information collected by sensors. They estimated heart rate independently of other signal features and designed an observation mechanism using known steady. This helps combine data from different wearable devices and signal patterns to improve monitoring accuracy. The performance of the proposed method is verified on real moving objects induced by ECG and Photoplethysmogram (PPG) data, and the results show an average error level lower than 2 times per minute.

To ameliorate the accuracy of emotion recognition, the authors of [44] have proposed a method using the fusion of physiological signals, they are subsampled to obtain the same number of images and the same sampling rate in order to be able to compare them, the next step is to create a new signal from the fusion of each of them in order to keep the essence of the signal. The idea is to fuse each signal, from each subject, for each emotion. They trained a feedforward Neural Network classifier based on the fused and unfused signals. Compared to existing methods, this method shows better performance on the two public datasets BP4D+ and DEAP.

A method based on Deep Learning has been presented in [45] allowing the multimodal data fusion to track patients' actions and activities in the execution of tasks requiring multiple steps by combining different modalities. The study consists in applying the proposed method to the fusion of an egocentric video from data extracted from a wearable wrist sensor on insulin self-injection. Based on a deep convolutional network (CNN), they extract visual features from individual video frames using a deep temporal fusion scheme. Then, a long and short-term memory (LSTM) network is applied to aggregate these features over time across various images. The complexity of the fusion stage is $O(N^2)$ where N depends on various parameters such as the length of the input video and motion features, the number of classes in the classification task, the number of memory blocks in the hidden layers for the video, motion, fusion sections and the number of connections in the network. The proposed method achieves results with 90% accuracy.

In the paper [46], the authors propose an updated Impulse Coupled Neural Network (PCNN) model that uses the Quantum Behavioral Particle Swarm Optimization (QPSO) algorithm to solve the PCNN parameter problem and ameliorate the efficiency and accuracy of the fusion of medical image. They used the following evaluation criteria : image entropy (EN) [47] which gives the amount of information in an image, average gradient (AG) which tells how clear the image is, and spatial frequency (SF) [48] that tells how active the image space is. The QPSO-PCNN algorithm is composed of six steps namely the initialization of the value of each particle, the local optimum and the global optimum, the evaluation of the initial population, the update of the particle positions, the update of the best positions, the verification of the final state and the output of the best results. The results obtained by this algorithm showed an improvement over the other methods.

The authors of [49] presented, based on Fuzzy Transformation (FTR), an algorithm for multimodal image fusion in the health domain by generating informative fused images taking into account the loss of data contained in erroneous images acquired through FTR that allows preserving the image features. This algorithm undergoes two types of evaluation, subjective and objective, the first is based on the fact that the merged images are judged by humans, this type of evaluation is generally expensive and time consuming. The second type is based on the following performance measures: Edge strength (Q) [50], Entropy (H) [51], Standard deviation (SD), Spatial frequency (SF), Feature mutual information (FMI), Fusion Factor (FF) [52], Structural similarity index measure (SSIM) [53]. Compared to other multimodal image fusion algorithms, this algorithm showed improvement in visual quality, mutual feature information, edge strength, texture and image clarity.

To categorize human sleep postures in a critical care atmosphere, the authors of [54] presented a formulation to merge two features derived from three different modalities, these features are extracted to develop a unimodal classifier and create a confidence level estimate. This method allows extracting features inde-

pendently, and then using them to estimate labels, unlike methods extracting a feature by fusing data that comes from different sensors.

In order to monitor patients in the IoT model based on smart intensive care units (ICU) using sensors, the authors of [55] gave the design of a health system, enabling monitoring to diagnose, prevent and control vital parameters of patients, especially in special cases where patients need monitoring to control their health and alert the medical body in real-time based on sensors connected to the ICU monitor and bonded to the patient's body. These sensors send signals to the monitors, which in turn display them and alert staff in case of an emergency. The Microsoft XBOX Kinect™ device was trained to contain sensors to collect data without forcing patients to attach sensors to themselves. This method was found to be 99.4% accurate.

An energy-efficient smart health system was proposed in [56], which can detect epileptic seizures based on fuzzy classification. Using a wireless EEG monitoring system, the EEG data collected by the edge devices to be processed is transferred to the Mobile Health Cloud (MHC) for further processing. This technique not only detects people suffering from seizures, but also people who are susceptible to them. The database used in this study was divided into three classes : seizure-free (SF) includes individuals with no history of seizures, non-active (NAC) includes people who are diagnosed with epileptic disorder while they are in non-active state, and active (AC) includes people who have epileptic seizure, in this case, an emergency notification will be sent to alert the staff. After the features are extracted, they are transformed into a fuzzy binary relation, in order to find the optimal rectangles to know the class to which the patients belong, and to build a classifier to detect the patient's condition. This method increased the battery life by 60% and gives accuracy of over 98%.

The authors of [57] proposed a system to monitor and detect voice disorders, this method introduces components such as assessors, patients, SLPs, service providers and the network structure, which is composed of edge computing to provide resources close to the user, and near real-time response, and cloud computing to provide significant processing and storage capacity. The collected samples are to the edge, then transferred them to the cloud, and afterwards sent them to the medical staff to prescribe the appropriate treatments for the patients. Two experiments were carried out, the detection of voice disorders which allows to know if the sample consists of people with voice disorders or not, and the classification of voice disorders which allows classifying the type of disorder, and, from these samples, the types of disorders selected are which or the number of people is high This technique showed an improvement in performance with an accuracy of over 98%.

Based on the research in [58], IoMT in healthcare is divided into four areas, We present a comparison in terms of: the task at hand, the IoMT tools used for each task, the classifier that represents the proposed algorithm, and finally the accuracy of each method. Table 2 shows IoMT and medical signals, Table 3 present IoMT and medical signals fusion, and Table 4 present Edge-Intelligent and Cloud-Based Smart HealthCare.

3 Discussion

According to Sect. 2, all of the work mentioned necessarily relies on the use of sensors and wearable devices to gather the information needed for the experiment presented. Each of these works is considered a revolution in the medical field, either through the use of technologies that no one has used before, or through the improvement provided by the algorithms.

Blood pressure, emotions, heart disease, sleep apnea, mental stress levels, brain tumors, heart rate, and seizures are some of the tasks that are addressed, they rely on IoMT tools namely EEC, EEG, MRI and others available in all medical institutes to measure vital signs, as well as devices such as cell phones, tablets, smartwatches, etc. Connected via the Internet to simplify the interaction between medical staff and patients, especially in critical cases. These tasks rely on the application of various artificial intelligence classifiers and edge-cloud computing, which have shown improvement over previous cloud-only methods. The application of AI algorithms such as CNN which is usually applied to recognize medical images and videos taken by different intelligent sensors, SVM which is used to solve discrimination problems in order to detect diseases, Neural Network which allows the analysis of electronic signals medical images and radiology [59], the Random Forest which allows making successive decisions thanks to a decision tree, LSTM which is useful for handwriting recognition [60], speech recognition [61,62], robot control [63], and others. Combining these technologies offers the benefits of each, but also poses major problems by combining their drawbacks, sometimes creating additional challenges that do not exist in any of them. For example, using AI algorithms with big data makes the analysis of information more complex, increasing the complexity of the algorithms and lengthening the processing time. However, in healthcare, time is considered an important factor in making critical life-saving decisions. The use of the Edge-Cloud opens the door to security issues, especially for massive data that will be transferred to the Cloud using architectures and protocols known by everyone, which increases the risk of breaking the integrity, confidentiality and availability of data and systems. In addition, the huge amount of data requires high bandwidth to ensure the transfer of patient information, which means a very high cost and also requires the use of gateways based on fog computing in order to connect different architectures especially as the medical field requires various health institutes to communicate with each other. Complexity is also not discussed in any of these papers. However, highly complex AI algorithms pose storage and processing resource challenges and increase processing time, especially as real-time response is important. And they increase power consumption, which means more money to spend.

Table 2. Comparison of various metrics related to IoMT and medical signals

Reference	Task	IoMT	Classifier	Accuracy
Fen Miao et al. [36]	Blood pressure measurement without cuff	Single channel ECG, 2 channel pulse wave sensors	Multi-instance regression algorithm	Estimated error 1.50
Fu Yang et al. [37]	Emotion recognition	32-channel EEG signals, 4-channel EOG signals, 4-channel EMG signals and vital signals	SVM	72% accuracy on DEAP and 89% accuracy on SEED
Qinghua Gu et al. [38]	Health and risk assessment for minors	Multi-sensors	SVM based on RF	RMSE < 0.2 and TSQ < 1.691 after 200 iterations
Muham-mad Muzamma et al. [39]	Heart disease prediction	Multi-sensors	RF, Kernel RF	98%
Tom Van Steenkiste et al. [40]	Prediction of sleep apnea	Multi-sensors	CNN, DL models LSTM	AUPR = 0.67 for CNN and AUPR = 0.71 for LSTM
Fares Al-Shargie et al. [41]	Mental stress rate	7 fNIRS channels were merged with 7 EEG electrodes	SVM	98%
Iv an Cabria et al. [42]	Brain tumor detection	MRIs	Intersection and union	Avg = 0.62 Std = 0.211 Med = 0.68
Viswam Nathan et al. [43]	Estimation of Heart Rate	Wearable sensors, ECG and PPG	Filtrage of Particules	Error < 2 beats/min Execution time:1.04 s (25 Hz sampling rate) and 1.18 s(125 Hz sampling rate)
Diego Fabiano et al. [44]	Emotion recognition	Physiological signals	Feedforward Neural Network	DEAP: 95.81% and BP4DC: 91.51%

Table 3. Comparison of various metrics related to IoMT and Medical Signals Fusion

Reference	Task	IoMT	Classifier	Accuracy
Edgar A Bernal et al. [45]	Human actions Recognition and activities for health surveillance	Wrist sensor, google glass wearable cameras	CNN and LSTM	90%
Xinzhe-ng Xu et al. [46]	Similarity measurements of different CT scan images	CT-MR, CT-MR T2, MR T1-MR T2, MR T1-FDG, MR T2-SPET images	QPSO-PCNN, PCNN	Entropy < 5.5
Y-D Ma et al. [49]	Generate a fused medical image from error images	MRI/CT, MRI T1/MRI T2, CT/PET, MRI/SPECT	Fuzzy transform	Fusion factor= 5.946 SSIM= 0.8871 FSIM= 0.8581
Yi LI et al. [54]	Sleep pose recognition	A carmine camera, RGB and depth sensors	Linear Discriminant Analysis (LDA), SVM	100% in bright scenes, 70% in poorly lit scenes, 90% in dark scenes

Table 4. Comparison of various metrics related to Edge-Intelligent and Cloud-Based Smart Health Care

Reference	Task	IoMT	Classifier	Accuracy
Iuliana Chiuchisan et al. [55]	Seizure detection	EEG	Kriging method	99.4%
Alaa Awad et al. [56]	Seizure detection	Wearable EEG	Swift In-Network Classification (SIC)	98%
Ghulam et al. [57]	Voice disorder	Smart sensors	CNN	98.5%

Most of the proposed methods have other limitations. In [36], the authors cited two main limitations regarding the proposed method, the scenarios are controlled and require the patients to remain relaxed which limits their movements, as well as the application of sensors to measure non-obstructive blood pressure is limited due to the current configuration. In [37], the features considered are generally common and do not support emotion recognition effectively. The method proposed in [39] requires significant computation because of its complexity. In [43], the authors' proposal is not addapted to patients with cardiac problems and they did not consider deployment and computational performance. The algorithm in [44] is applied just on a specific number and not on the whole database, so it may not be applicable on a large number of data. In [45], the

methods applied are deep because of the different statistical characteristics used and cannot be combined with shallow models. In [46], the sharpness of the fused images still remains a challenge. The energy and delay savings achieved by the system in [56] are based on specific characteristics of the data.

4 Challenges and Future Directions

A large-scale healthcare system will need to manage a large number of patients treated in a hospital, which means that a large amount of data collected from heterogeneous sensors must be processed, analyzed and stored. This poses challenges in terms of patient information processing and security due to the increasing complexity of the data and the rise of cyber attacks [64], especially with the use of the cloud. Edge computing can reduce these challenges due to its distributive nature, storing data in different locations and not aggregating it in the Cloud [3], but this is still an ineffective solution for ensuring the confidentiality and privacy of patient medical data. Denial of service attacks must be taken into account, especially since replication is not always an option, as many devices are considered as implementations of base systems [58], as well as data replication increases the need for storage resources which implies huge costs, due to the huge amount of medical data collected.

The effective use of AI is also considered a challenge, as remote monitoring of patients using wearable sensors can result in lost or incomplete data before being recorded in the cloud [64].

Managing and controlling IoT devices to diagnose diseases, and data collected by different sensors poses hardware issues, discharged batteries or connectivity issues, loss of electrical power can cause data loss [64], as well as risks to patients' lives such as toxicity and ignition risk [3].

In our next work, we are looking for an optimal architecture based on different IoMT technologies combined with cloud, edge and fog computing, capable of minimizing these challenges, both in terms of hardware and applications, taking into account the security and complexity of the AI algorithms used, which will lead us to reduce costs.

5 Conclusion

Smart healthcare is an important field that allows, through the use of IoMT, cloud, edge, AI and wearable devices, to connect different medical institutes with each other, as well as staff and patients, making it easier to meet the needs of patients, especially in critical cases as well as in cases that require regular follow-up. This paper presents the different architectures and technologies that exist to improve the service of intelligent healthcare, and the applications of these technologies in cases of diseases, either for detection and diagnosis, or for decision-making to treat them. It also presents a comparison between existing work in terms of IoMT technology, AI algorithms and fusion algorithms used for different tasks. Then, shows a comparison between various papers on tasks, IoMT

tools, classifier, accuracy, and complexity levels. The use of fusion algorithms has become necessary, they enable easier treatment and diagnosis through sensors and wearable devices attached to the patient's body. So that all important features can be collected using IoMT tools. However, most of these technologies do not fully guarantee the security and privacy of patient data and suffer from high energy costs.

References

1. Tian, S., Yang, W., Grange, J.M.L., Wang, P., Huang, W., Ye, Z.: Smart healthcare: making medical care more intelligent. *Glob. Health J.* **3**(3), 62–65 (2019)
2. Yadav, R., et al.: Smart healthcare: RL-based task offloading scheme for edge-enable sensor networks. *IEEE Sens. J.* **21**(22), 24910–24918 (2021)
3. Hartmann, M., Hashmi, U.S., Imran, A.: Edge computing in smart health care systems: review, challenges, and research directions. *Trans. Emerg. Telecommun. Technol.* **33**(3), e3710 (2022)
4. Bhunia, S.S.: Sensor-cloud: enabling remote health-care services. In: *Proceedings of the 2015 on MobiSys PhD Forum*, pp. 3–4 (2015)
5. Althebyan, Q., Yaseen, Q., Jararweh, Y., Al-Ayyoub, M.: Cloud support for large scale e-healthcare systems. *Ann. Telecommun.* **71**(9), 503–515 (2016)
6. Klonoff, D.C.: Fog computing and edge computing architectures for processing data from diabetes devices connected to the medical internet of things. *J. Diab. Sci. Technol.* **11**(4), 647–652 (2017)
7. Xiao, Y., Jia, Y., Liu, C., Cheng, X., Jiguo, Yu., Lv, W.: Edge computing security: state of the art and challenges. *Proc. IEEE* **107**(8), 1608–1631 (2019)
8. Razdan, S., Sharma, S.: Internet of medical things (iomt): overview, emerging technologies, and case studies. *IETE Tech. Rev.* **39**, 1–14 (2021)
9. Sun, L., Jiang, X., Ren, H., Guo, Y.: Edge-cloud computing and artificial intelligence in internet of medical things: architecture, technology and application. *IEEE Access* **8**, 101079–101092 (2020)
10. Das, P.K., Zhu, F., Chen, S., Luo, C., Ranjan, P., Xiong, G.: Smart medical healthcare of internet of medical things (iomt): application of non-contact sensing. In: *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 375–380. IEEE (2019)
11. Elaiyaraaja, K., Kumar, M.S., Chidhambararajan, B.: Deep learning-based bdmsf resource sharing-a systematic approach for analysis and visualization. In: *Disruptive Technologies for Big Data and Cloud Applications*, pp. 75–84. Springer, Heidelberg (2022). https://doi.org/10.1007/978-981-19-2177-3_8
12. Lamonaca, F., et al.: A new measurement system to boost the iomt for the blood pressure monitoring. In: *2019 IEEE International Symposium on Measurements & Networking (M&N)*, pp. 1–6. IEEE (2019)
13. Lamonaca, F., et al.: An overview on internet of medical things in blood pressure monitoring. In: *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1–6. IEEE (2019)
14. Young, W., Corbett, J., Gerber, M.S., Patek, S., Feng, L.: Damon: a data authenticity monitoring system for diabetes management. In: *2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pp. 25–36. IEEE (2018)

15. Liu, J., Tang, H., Sun, R., Xiaojiang, D., Guizani, M.: Lightweight and privacy-preserving medical services access for healthcare cloud. *IEEE Access* **7**, 106951–106961 (2019)
16. Ning, Z., Huang, J., Wang, X., Rodrigues, J.J.P.C., Guo, L.: Mobile edge computing-enabled internet of vehicles: toward energy-efficient scheduling. *IEEE Netw.* **33**(5), 198–205 (2019)
17. Chen, H., Zhao, T., Li, C., Guo, Y.: Green internet of vehicles: architecture, enabling technologies, and applications. *IEEE Access* **7**, 179185–179198 (2019)
18. Wang, X., Ning, Z., Wang, L.: Offloading in internet of vehicles: a fog-enabled real-time traffic management system. *IEEE Trans. Ind. Inf.* **14**(10), 4568–4578 (2018)
19. Ning, Z., Huang, J., Wang, X.: Vehicular fog computing: enabling real-time traffic management for smart cities. *IEEE Wirel. Commun.* **26**(1), 87–93 (2019)
20. Pathinarupothi, R.K., Durga, P., Rangan, E.K.: Iot-based smart edge for global health: remote monitoring with severity detection and alerts transmission. *IEEE Internet Things J.* **6**(2), 2449–2462 (2018)
21. Queralta, J.P., Gia, T.N., Tenhunen, H., Westerlund, T.: Edge-ai in lora-based health monitoring: Fall detection system with fog computing and lstm recurrent neural networks. In: 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), pp. 601–604. IEEE (2019)
22. Emam, A., Abdellatif, A.A., Mohamed, A., Harras, K.A.: Edgehealth: an energy-efficient edge-based remote mhealth monitoring system. In: 2019 IEEE wireless communications and networking conference (WCNC), pp. 1–7. IEEE (2019)
23. Esmaeilzadeh, P.: Use of ai-based tools for healthcare purposes: a survey study from consumers' perspectives. *BMC Med. Inf. Decis. Mak.* **20**(1), 1–19 (2020)
24. Väänänen, A., Haataja, K., Vehviläinen-Julkunen, K., Toivanen, P.: Ai in health-care: a narrative review. *F1000Research* **10**(6), 6 (2021)
25. Stuart, J., et al.: Artificial intelligence a modern approach third edition (2010)
26. Tucker, A.B.: Computer Science Handbook. Chapman and Hall/CRC, Boca Raton (2004)
27. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
28. Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.T.: Deep learning for health-care: review, opportunities and challenges. *Brief. Bioinf.* **19**(6), 1236–1246 (2018)
29. Drake, M.: Encyclopedia of Library and Information Science, vol. 1. CRC Press, Boca Raton (2003)
30. Chowdhury, G.C.: Natural language processing dept. of computer and information sciences university of strathclyde. Glasgow G1 1XH, UK (2003)
31. Chawla, N.: Ai, iot and wearable technology for smart healthcare? a review. *Int. J. Green Energy* **7**(1), 9–13 (2020)
32. Adarsha, A.S., Reader, K., Erban, S.: User experience, iomt, and healthcare. *AIS Trans. Human-Comput. Interact.* **11**(4), 264–273 (2019)
33. Islam, S.M.R., Kwak, D., Kabir, M.D.H., Hossain, M., Kwak, K.S.: The internet of things for health care: a comprehensive survey. *IEEE Access* **3**, 678–708 (2015)
34. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., Ayyash, M.: Enabling technology for the internet of health things. *IEEE Transl. Content Mining* **17**, 2347–2376 (2015)
35. Iroju, O.G., Olaleke, J.O.: A systematic review of natural language processing in healthcare. *Int. J. Inf. Technol. Comput. Sci.* **8**, 44–50 (2015)
36. Miao, F., Liu, Z.-D., Liu, J.-K., Wen, B., He, Q.-Y., Li, Y.: Multi-sensor fusion approach for cuff-less blood pressure measurement. *IEEE J. Biomed. Health Inf.* **24**(1), 79–91 (2019)

37. Yang, F., Zhao, X., Jiang, W., Gao, P., Liu, G.: Multi-method fusion of cross-subject emotion recognition based on high-dimensional eeg features. *Front. Comput. Neurosci.* **13**, 53 (2019)
38. Qinghua, G., Jiang, S., Lian, M., Caiwu, L.: Health and safety situation awareness model and emergency management based on multi-sensor signal fusion. *IEEE Access* **7**, 958–968 (2018)
39. Muzammal, M., Talat, R., Sodhro, A.H., Pirbhulal, S.: A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. *Inf. Fusion* **53**, 155–164 (2020)
40. Van Steenkiste, T., Deschrijver, D., Dhaene, T.: Sensor fusion using backward shortcut connections for sleep apnea detection in multi-modal data. In: *Machine Learning for Health Workshop*, pp. 112–125. PMLR (2020)
41. Al-Shargie, F.: Fusion of fnirs and eeg signals: Mental stress study (2019)
42. Cabria, I., Gondra, I.: MRI segmentation fusion for brain tumor detection. *Inf. Fusion* **36**, 1–9 (2017)
43. Nathan, V., Jafari, R.: Particle filtering and sensor fusion for robust heart rate monitoring using wearable sensors. *IEEE J. Biomed. Health Inf.* **22**(6), 1834–1846 (2017)
44. Fabiano, D., Canavan, S.: Emotion recognition using fused physiological signals. In: *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 42–48. IEEE (2019)
45. Bernal, E.A., et al.: Deep temporal multimodal fusion for medical procedure monitoring using wearable sensors. *IEEE Trans. Multimedia* **20**(1), 107–118 (2017)
46. Xinzheng, X., Shan, D., Wang, G., Jiang, X.: Multimodal medical image fusion using pcnn optimized by the qpso algorithm. *Appl. Soft Comput.* **46**, 588–595 (2016)
47. Ma, Y.-D., Dai, R.-L., Li, L.: Automated image segmentation using pulse coupled neural networks and image's entropy. *J.-China Inst. Commun.* **23**(1), 46–51 (2002)
48. Yi, L.I., Xiao-jun, W.U.: A novel image fusion method using self-adaptive dual-channel pulse coupled neural networks based on pso evolutionary learning. *Acta Electronica Sinica* **42**(2), 217 (2014)
49. Manchanda, M., Sharma, R.: An improved multimodal medical image fusion algorithm based on fuzzy transform. *J. Visual Commun. Image Represent.* **51**, 76–94 (2018)
50. Xydeas, C.S., Petrovic, V., et al.: Objective image fusion performance measure. *Electron. Lett.* **36**(4), 308–309 (2000)
51. Ben Hamza, A.: Jensen-rnyi divergence measure: theoretical and computational perspectives. In: *IEEE International Symposium on Information Theory* (2003)
52. Arathi ,T., Soman, K.P.: Performance evaluation of information theoretic image fusion metrics over quantitative metrics. In: *2009 International Conference on Advances in Recent Technologies in Communication and Computing*, pp. 225–227. IEEE (2009)
53. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
54. Torres, C., Hammond, S.D., Fried, J.C., Manjunath, B.S.: Sleep pose recognition in an ICU using multimodal data and environmental feedback. In: Nalpantidis, L., Krüger, V., Eklundh, J.-O., Gasteratos, A. (eds.) *ICVS 2015. LNCS*, vol. 9163, pp. 56–66. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20904-3_6

55. Chiuchisan, I., Costin, H.N., Geman, O.: Adopting the internet of things technologies in health care systems. In: 2014 International Conference and Exposition on Electrical and Power Engineering (EPE), pp. 532–535. IEEE (2014)
56. Abdellatif, A.A., Emam, A., Chiasserini, C.F., Mohamed, A., Jaoua, A., Ward, R.: Edge-based compression and classification for smart healthcare systems: concept, implementation and evaluation. *Expert Syst. Appl.* **117**, 1–14 (2019)
57. Muhammad, G., Alhamid, M.F., Alsulaiman, M., Gupta, B.: Edge computing with cloud for voice disorder assessment and treatment. *IEEE Commun. Maga.* **56**(4), 60–65 (2018)
58. Alshehri, F., Muhammad, G.: A comprehensive survey of the internet of things (iot) and ai-based smart healthcare. *IEEE Access* **9**, 3660–3678 (2020)
59. Patel, J.L., Goyal, R.K.: Applications of artificial neural networks in medical science. *Curr. Clin. Pharmacol.* **2**(3), 217–226 (2007)
60. Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., Schmidhuber, J.: A novel connectionist system for unconstrained handwriting recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(5), 855–868 (2008)
61. Sak, H., Senior, A.W., Beaufays, F.: Long short-term memory recurrent neural network architectures for large scale acoustic modeling (2014)
62. Li, X., Wu, X.: Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4520–4524. IEEE (2015)
63. Mayer, H., Gomez, F., Wierstra, D., Nagy, I., Knoll, A., Schmidhuber, J.: A system for robotic heart surgery that learns to tie knots using recurrent neural networks. *Adv. Rob.* **22**(13–14), 1521–1537 (2008)
64. Nasr, M., Islam, M.M., Shehata, S., Karray, F., Quintana, Y.: Smart healthcare in the age of ai: recent advances, challenges, and future prospects. *IEEE Access* **9**, 145248–145270 (2021)



The Impact of Leadership on Employee Motivation in the Jordanian Telecommunication Sector

Haron Ismail Al-lawam¹, Mohammad Salameh Zaid Almahairah²,
Hiba Hussein Mohammad Almomani³ , Ahmad A. I. Shajrawi⁴,
Saddam Rateb Darawsheh⁵ , Anwar Saud Al-Shaar⁶,
and Muhammad Turki Alshurideh⁷

¹ Universiti Malaysia Terengganu, Kuala Terengganu, Terengganu, Malaysia

² Department: Management Information System, Asra University Jordan, Amman, Jordan
mohammad.almahirah@iu.edu.jo

³ Faculty of Educational Sciences, Isra University, Amman, Jordan
hiba.almomani@iu.edu.jo

⁴ Ministry of Education, Amman, Jordan

⁵ Department of Administrative Sciences, The Applied College, Imam Abdurrahman Bin Faisal University, P.O. Box: 1982, Dammam, Saudi Arabia
srdarawsehe@iau.edu.sa

⁶ Amman, Jordan

⁷ Department of Marketing, School of Business, The University of Jordan,
Amman 11942, Jordan
m.alshurideh@ju.edu.jo

Abstract. The main objective of the current study is to examine the relationship between leadership style and employee motivations from the point of view of motivation in the Jordanian telecommunication sector the population is 70 the appropriate sample size will be 59, (Telecommunication Regulatory Commission 2023), therefore the sample of this study is 59 directors, a total of 59 questionnaires were distributed to the respondent's development model by (Bass et al. 2003) employee motivation Scale, results showed it appears that transformational leadership and transactional leadership was correlated with employee motivation, but laissez-faire leadership has no significant relationship with employee motivation.

Keywords: Leadership Styles · Transformational leadership · Transactional leadership · Laissez faire leadership · Employee motivation

1 Introduction

It is important for employers to create a motivating and positive environment and for employees to increase their job satisfaction, (Alameeri et al. 2020; Al Kurdi et al. 2020), when they understand the group's mission and encourage peers to think beyond their own selfishness for the benefit of the group (Bass et al. 2003; Nuseir et al. 2021),

they are much like a charismatic leader, not only encourages employees to demonstrate leadership, but also gives employees the freedom to think and work to improve the performance, strategy, and structure of the organization, and organization, fostering product and technology innovation within the organization. Motivate and inspire (Bass and Stogdill 1990; Alshurideh et al. 2022). It has been shown to produce positive and significant improvements in employee attitudes, happiness, civic conduct, dedication, effectiveness, effort, and financial performance are more effective than other leadership styles (Odeh et al. 2021; Allozi et al. 2022). In particular, they seem to motivate their followers through emotion (Church 2012; Alzoubi et al. 2021). There are many studies on leadership and the relationship between the characteristics of the leader and motivation with the behavior of the employee, leadership is a process through which the executive officer can direct and influence the behavior and action of employees towards achieving specific goals in a specific situation, motivation is a word derived from the word 'motive' that Means needs and desires or drive within individuals, it is the process of motivating people to take action to achieve goals in the organisation.

Leadership can be shown in different ways in different situations, most of his people think that a leader is simply a district president or prime minister, or just a political leader, but leadership is not limited to just political leaders, mothers and fathers show leadership by raising, health workers take the lead in developing services that meet the needs of the communities they serve, or work with other organizations to create low-cost, prevention-focused programs and services; to exercise leadership, protect people and take over the security of the place, team his leader and so on (Al-Dhuhouri et al. 2020; Al Khayyal et al. 2020; AlShehhi et al. 2020; Kurdi et al. 2020; Harahsheh et al. 2021). In his society today, leader's primary task is to motivate and retain employees, motivation is the lifeblood of a company, a well-managed company is able to motivate and retain employees, giving it a competitive advantage: Less volatility, increased productivity for the less absenteeism (Obeidat et al. 2021; Abuanzeh et al. 2022). Increasing sales and improving performance is the combination of basic knowledge, skills, and availability of resources that influence employee attitudes toward the achievement of organizational goals. (Amin et al. 2013; Alsuwaidi et al. 2020; Al Kurdi et al. 2022). Transformational leadership has been a buzzword for decades, but as the challenges of the Industry approach, it has evolved into a powerful term that unlocks employee skills through effective communication networks, awareness, and the need for accepted, we live up to the expectations of our employees and act in line with our organization's vision and mission, value is what counts, Headers are unconsciously moral and ethical, but followers learn and imitate every day, a reputation built by a leader boosts morale and confidence, because the ethical and moral behavior of leaders, and their honesty and fairness to followers, actually help motivate them to work harder, Promote the development and utilization of transformational leadership creates the flexibility to grow in times of change; transformational leadership develops the skills necessary to contribute to better career growth and advancement, these leaders are equipped with the ability to build excellent communication networks, positive work environments, and deeper understandings that help build greater energy, enthusiasm, and passion, The aim of conducting this research is to find out the impact of leadership on motivating employees for a brighter future and growth for the organization.

Telecom in Jordan is characterized by dynamic competition and growing growth and is constantly modernized and expanded, it happens to be the most competitive in the entire region, the telecommunications sector is one of the most important economic sectors in Jordan, and communications can be seen as a gateway to Jordan's economic growth and social life through the development of services (Enas et al. 2019), the government of Jordan is mobilizing the most extensive possible access to communication services at reasonable costs, and is ensuring that the telecommunication sector is fair game in terms of competition (Ashour 2014), the industry faces a need for increased employee motivation in the current scenario of high competition, high career growth requirements, and high employee demand for better bargaining positions compared to the sheer number of employees, in order to meet the needs of the current workforce's increased motivation needs to be met, option, Positive effects on motivation has been elucidated, future recommendations are also made to meet employee needs and expectations in order to retain employees, the higher the motivation, the higher the level of performance, organizations require leaders and employees to work together on a variety of tasks to achieve desired results, every leader should have her own Leadership style, (Molero 2007), state that in the era of leadership, leaders need to think and demonstrate leadership styles, through certain leadership styles, employers can influence the performance of people within their organization, in other words, leadership style plays an important role in running an organization.

In addition to management style, an important value element of an organization is motivated to work, that work motivation is an individual's environment that fosters the willingness to perform specific activities to achieve goals, motivation is a factor that leads to jobs at is a faction in and improves performance in employees, in organizational life, employees need a boost as a form of work motivation, to improve their performance, Work motives must therefore be given serious attention by leaders or managers in organizations, especially those whose daily work is in direct contact with their subordinates at work. Motivation is therefore a key variable that organizations must seriously consider in order to improve employee performance.

2 Literature Review

Leadership is the process by which one people influence a group of people to go in a particular direction, the Leadership has been used in many aspects of human endeavor, including politics, economics, science, and social enterprise, company or organization leaders also play a key role in the performance of followers and their organizations by motivating them (Gopal 2014), leaders are key figures in the success or failure of an institution as stated in a study conducted by (Bass 1990), the total success or failure factors are based on leaders, manager style influences employee behavior and acceptance of Transformational Leadership and its impact on employee motivation the cause of this take a look at is to fulfill the organization's intention of higher destiny and boom the organization wants to increase its production and development by motivating employees and better career desires and prosperity, in addition to better communication for employees, the wonderful effects of motivation have been clarified, and future indicators have been developed on the way to satisfy the desires and expectations of the team of workers

in order to retain them and increase their productivity, The better the motivation, the better may be the overall performance levelsthe function of leader's non-public attributes and behavior of a transformational management in increasing at the goal and reason of any worker in adapting to the organizational change, there is a tremendous relation in principal's variation and transition to the change, the moderation among the family members of teacher's resistance and variation has been defined to make contributions to the literature. Further, Oreg and Berson (2011) suggested considering innovative leadership styles such as democratic, autocratic, or task- or relationship-oriented leadership. Innovative leadership and other leadership styles have been found by to lead to improved employee satisfaction and productivity using a sample of 147 respondents working in 35 work teams, transformational leadership style, relational association, or democratic or task-related leadership style a significant increase in the relationship between leadership style and organization yields his-oriented variables attributed to subordinates' performance, satisfaction, and additional effort (Gautama and Mishra 2020).

A leader is someone who motivates employees to achieve a common goal, there are also many theories and leadership styles, but different sectors or organizations may require specific leadership styles, and effective leaders need to know which leadership styles fit their employees' motivational needs, it is impossible for an organization to succeed without a good leader (Hafeez and Mariam 2012), Leaders must discover strategies to motivate their employees, it should encourage and motivate employees to implement, The tasks of the organization and its development and increase its productivity, leaders who have knowledge of various motivational theories that are more likely to select the appropriate theories for the appropriate situation and the employee, so this boss has subordinates who behave better and are happy as a result of the leader's style of motivation (Hanifa and Setiawan 2014), (Hughes 2015; Hussain 2012), Leaders need to understand their employees and their motivations a high level of motivation makes a very important contribution to organizational performance motivated employees produce at the highest possible level and work harder than unmotivated employees characteristics of motivated employees being part of a team at work they are interested in helping others and support them in their work. They generally put more effort into their work in the organization, (Hussain and Ali 2012).

According to Burns, transformational management is an ethical workout in that it increases the usual human conduct, in assessment all charismatic leaders aren't transformational, they also are charismatic because of the articulation of a compelling imaginative, and prescient Impact of Leadership on Employee motivation telecommunication Sector destiny and shape sturdy emotional attachments with the subordinates, transformational management regularly engage with troubles and change, consequently transformational chief have to be geared up to stand conflict, transformational leadership has been continually claimed to be a greater power than the leadership styles, (Jerotich 2013).

He explained that transformative leaders' directive or participatory, authoritarian or democratic, trends and inspire change supporting and participating in open consultations can be participatory and transformative, it can be directive and transactional in promising, and transactional in reaching common power-sharing agreements, the demonstrates both transformative and transacting behavior, as well as mixed direction and participation. The

Transformation Guide focuses on Idealized Influence, Inspiring motivation, intellectual Stimulation, and individualized reflection, (Hussain and Ali 2012),

Transactional leaders try to meet the present-day wishes, and transactional leaders achieve agreed-upon objectives without encouraging them to tackle more obligations, There isn't any try to alternate fans' attitudes, values, growth, and improvement on a long-time period basis, both leader's and fans consciousness on accomplishing the negotiated overall performance level transactional chief motivates subordinates with the aid of using giving rewards for offerings provided, this chief clarifies the subordinates dreams and arranges contingent rewards as inducements towards the fulfillment of the objective, one may theorize transactional leaders could have the finest, (Joseph and Winston 2005).

Leadership behaviors that emphasize telling or controlling could be labeled as transactional management due to the fact that rewards and field are administered in steps transactional management is an approach requiring a consistent application, there are additives of transactional leadership and management with the aid of using the exception, contingent praise is when the chief offers rewards if the subordinate work according to the overall performance expectancies or expends the essential attempt ought to additionally relate definitely to overall performance, those leaders make clear expectancies and apprehend achievements that definitely make a contribution to better tiers of attempt and overall performance, (Joseph and Winston 2005).

Laissez-faire leaders Leadership avoidance such as "absent when necessary" 'and "not taking action when problems become chronic were strongly associated with subordinate dissatisfaction, conflict, and inefficiency, laissez-faire leadership is a passive type of leadership style that is seen as not caring about others, there is no relationship exchange between the leader and his followers, it describes non-transactional leadership in which necessary decisions are not made, action is delayed, leadership mandates are ignored, and authority is not used (Bass 1993), supply their group contributors with a number of motivations in how they do their work, and the way they set their objectives, they offer aid with sources and recommendations if needed, however in any other case they do not get involved, this autonomy can cause excessive activity work, however, it could be destructive if group contributors do not control their time well, or in the event that they do not have the knowledge, skills, or self-motivation to do effectively, laissez-faire management also can arise while managers do not have manage over their paintings and their people, lewis's frameworks famous and useful, as it encourages leaders to be increase product from motivation employees, (Rae 2008).

3 Role of Leadership Style in Employee Motivation

Leadership style is one of the most important factors in employee motivation. Employee responses to leaders depend on how employees create the environment (Chowdhury 2014), a change in management style allowed employees to experience a different environment, there is a strong correlation between managerial and employee motivation (Uddin 2019), when employees are unmotivated and not treated with an effective leadership approach that is appropriate to the work environment, work culture and organizational structure, the consequences are detrimental to the organization, leading to

dissatisfaction and the loss of valuable employees, an effective leadership approach is a key factor in creating a successful organization and in motivating employees by projecting a positive image into their minds, leadership approach is an important motivating factor, employee responses to leaders tend to depend on the characteristics of both employees and managers (Mengesha 2015). A leader who believes that others can take part in decision-making and keep them accountable, this is the idea for constructing the engagement of others and the ideals of these around them, whether or not they exercise management thru consensus or consensus, may be selected on the idea that it could contradict legal guidelines inside an organization, empower society with the aid of using converting the panorama of the regulation and increasing it in keeping with its guidelines. She is stimulated to paintings tough to make her personal goals come true, she became promised this form of supervisor values obligation, values collective participation, loves to examine others, and regularly has brilliant and revolutionary bosses who're employed to make their goals come true (Hussain and sinjar Alsamaray 2014), democratic leaders can take obligation for his or her errors and take delivery of the consequences, unlike autocratic bosses they're regularly above the regulation does management that believes in competence have the proper of others to take part in and be held responsible for the decision-making system it should constitute the dreams to be achieved, this bureaucracy the premise, with the aid of using desire or agreement, for others to construct round them the participation and ideals to take management positions.

Their packages and initiatives can be decided on the premise that they violate numerous legal guidelines winning with inside the organization or network and that they should follow the regulation to extrude the regulation and formulate rules to gain the promised dreams affects his fans with the aid of using the use of symbols tales and pics to gain awesome consequences democratic leaders have a imaginative and prescient for the destiny achievement in their supporters, moreover, leaders are inclined to take any dangers to illustrate their commitment, according to preceding research (Anwar and Qadir 2017), and others (Anwar 2017), management refers to an individual's capacity to steer different corporations of human beings to gain a aim or set of dreams consequently, management is all approximately motivating people to duplicate extrude and imaginative and prescient (Anwar and Louis 2017), we discovered a wonderful affiliation among democratic management and OCB Further discovered a dating among charismatic management and organizational achievement. Furthermore, (Anwar and Ghafoor 2017), discovered the wonderful institutions among charismatic management and his OCB, additionally, (Anwar and Balcioglu 2016).

4 Theoretical Framework and Hypotheses Development

Based on the literature review, the following hypothesis and framework were developed (Fig. 1).

- H1: Transactional leadership has a positive relationship with employee motivation in the Jordanian telecommunication Sector.
- H2: Transformational leadership has a positive relationship with employee motivation in the Jordanian telecommunication Sector.

- H3: Laissez-faire leadership has a positive relationship with employee motivation in the Jordanian Telecommunication Sector.

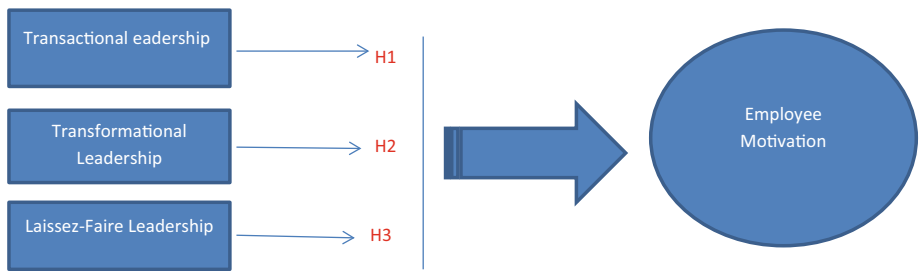


Fig. 1. Conceptual framework of the research

5 Research Methodology and Data Analysis

The number of directors of the Jordanian telecommunication Sector (<http://www.trc.gov.jo>), regarding this study, the krejcie and morgan table was used to determine the appropriate sample size for this study. According to Krejcie and Morgan (1970), if the population is 70 the appropriate sample size will be 59 (Telecommunication Regulatory Commission 2023), therefore the sample of this study is 59 directors, a total of 59 questionnaires were distributed to the respondents.

The multifactor leadership questionnaire (MLQ) and the WMS motivator, the MLQ model, were changed to a 5-factor scale for easier coding and data interpretation. MLQ was developed by Bass (1997) and covers a wide range of leadership types, from passive leaders to leaders who conditionally reward their followers to leaders who turn their followers into leaders themselves, it was adjusted to consist of 29 items that measure major characters, data analysis was performed using SPSS (Social Science Statistical Package) to establish validity, cronbach’s alpha’s employee motivation scale reliability statistic is 0.834 on 10 items, and Cronbach’s alpha’s leadership approach reliability statistic is 0.862 on 29 items.

Table (1) Shows reliability tests for three independent variables and one dependent variable. Innovative leadership style $\alpha = 0.834 > 0.7$ (Kothari 2004) Employee motivation $\alpha 0.862 > 0.7$. The 29 items are scored as follows: The 10-item motivational model was modified to a 7-factor scale for ease of coding and data interpretation, and the 10 items are scored as follows: Demographic items are based on previous theoretical and empirical studies (Table 1).

The correlation between transformational leadership style and employee engagement is positive, with a score of 0.602. The correlation between transaction style and motivation is also positive, with a score of 0.329. However, the correlation is low, implying that transformational leadership styles motivate employee’s more than transactional styles, whereas laissez-faire styles are negatively correlated with motivation, in other words, employees are unhappy with laissez-faire leadership.

Table 1. Pearson Product-moment correlations between independent and dependent variables

Variables	1	2	3	4
Employee Motivation	1			
Transformational leadership (IV)	.602	1		
Transactional leadership (IV)	.329	.845	1	
Laissez-Faire Leadership (IV)	-.585	-.732	-.496	1

6 Conclusion

The aim of this study was to investigate the relationships between leadership and motivation in Jordan's telecom sector this study showed that there is a relationship between transformational leadership motivations in the Jordanian telecommunication sector, there is a positive correlation between individual attention and all sub-variables of employee motivation, these results are consistent with previous studies showing that transformational leadership factors have a significant positive impact on motivation and laissez-faire leadership has a significant negative impact on subordinate motivation (Tazeem and Muhammad 2011), (Bass and Avolio 1994; Loke 2001; Bass 1998; Avolio 1999; Shim et al. 2002; Waldman et al.), this finding is consistent with several previous studies that have found that fulfilling contingent reward promises has a significant impact on employee motivation. Commentators always see rewards and encouragement as one of the most important motivations. (Snape 1996; Erkutlu 2008), managers have motivations such as discretionary aspects, (freedom to choose what, when, and how to perform activities), Particular attention should be paid to job requirements (controlled or uncontrolled activity rate), appropriate use of skills, stressors and competencies, Or laissez-faire - a style for consistently and efficiently increasing employee engagement, acquisitions are recommended because they improve employee performance quality and company performance, managers should avoid all laissez-faire practices and instead spend time coaching, paying attention to employee skills and needs, helping employees develop their talents, and promoting higher performance standards within the organization, should provide a supportive environment that helps we also need to increase our knowledge of how leadership styles affect employees, you should choose the style that best suits your company's goals and the needs and desires of your employees, to improve our employees, we must act as ethical role models.

Jordanian employee telecommunications are motivated by social rewards, the need for self-actualization, rewards, and improved working conditions. A leader with a compelling vision must "talk" to win the hearts and minds of employees. It is very important that the leaders of the organization always communicate the vision and make sure there is no doubt about the direction. Ken Blanchard, a world-renowned management coach, said, "Focusing on employee engagement and empowerment, engagement and discretion, role streamlining, financial success, customer satisfaction, employee performance drive, etc., all increase the motivation of the organization.

References

- Abuanzeh, A., Alnawayseh, A., Qtaishat, G., Alshurideh, M.: The role of strategic agility towards competitiveness with mediating effect of knowledge management. *Uncertain Supply Chain Manag.* **10**(4), 1523–1534 (2022)
- Gautam, A., Jain, P., Mishra, A.K.: Impact of transformational leadership on employee motivation. *Int. J. Manag.* (IJM) **11**(5), 2132–2137 (2020)
- Alameeri, K., Alshurideh, M., Al Kurdi, B., Salloum, S.A.: The effect of work environment happiness on employee leadership. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *International Conference on Advanced Intelligent Systems and Informatics*, pp. 668–680. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58669-0_60
- Al-Dhuhouri, F.S., Alshurideh, M., Al Kurdi, B., Salloum, S.A.: Enhancing our understanding of the relationship between leadership, team characteristics, emotional intelligence and their effect on team performance: a critical review. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *International Conference on Advanced Intelligent Systems and Informatics*, vol. 1261, pp. 644–655. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58669-0_58
- Al Khayyal, A.O., Alshurideh, M., Al Kurdi, B., Salloum, S.A.: Women empowerment in UAE: a systematic review. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *International Conference on Advanced Intelligent Systems and Informatics*, pp. 742–755. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58669-0_58
- Al Kurdi, B., Alshurideh, M., Al afaishat, T.: Employee retention and organizational performance: evidence from banking industry. *Manag. Sci. Lett.* **10**(16), 3981–3990 (2020)
- Al Kurdi, B., Alzoubi, H., Akour, I., Alshurideh, M.: The effect of blockchain and smart inventory system on supply chain performance: empirical evidence from retail industry. *Uncertain Supply Chain Manag.* **10**(4), 1111–1116 (2022)
- Allozi, A., Alshurideh, M., AlHamad, A., Al Kurdi, B.: Impact of transformational leadership on the job satisfaction with the moderating role of organizational commitment: case of UAE and Jordan manufacturing companies. *Acad. Strateg. Manag. J.* **21**, 1–13 (2022)
- Alnsour, M., Abu Tayehamli, B., Awwad Alzyadat M.: Using SERVQUAL to assess the quality of service lrovided by Jordanian telecommunications sector. *Int. J. Commer. Manag.* **24**(3) 209–218 (2014)
- AlShehhi, H., Alshurideh, M., Al Kurdi, B., Salloum, S.A.: The impact of ethical leadership on employees performance: a systematic review. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *International Conference on Advanced Intelligent Systems and Informatics*, pp. 417–426. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58669-0_38
- Alshurideh, M., Kurdi, B., Alzoubi, H., Obeidat, B., Hamadneh, S., Ahmad, A.: The influence of supply chain partners' integrations on organizational performance: the moderating role of trust. *Uncertain Supply Chain Manag.* **10**(4), 1191–1202 (2022)
- Alsuwaidi, M., Alshurideh, M., Al Kurdi, B., Salloum, S.A.: Performance appraisal on employees' motivation: a comprehensive analysis. In: Hassanien, A.E., Slowik, A., Snášel, V., El-Deeb, H., Tolba, F.M. (eds.) *International Conference on Advanced Intelligent Systems and Informatics*, vol. 1261, pp. 681–693. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58669-0_61
- Alzoubi, H., Alshurideh, M., Akour, I., Shishan, F., Aziz, R., Al Kurdi, B.: Adaptive intelligence and emotional intelligence as the new determinant of success in organizations. An empirical study in Dubai's real estate. *J. Legal Ethical Regul. Issues* **24**(6), 1–15 (2021)
- Amin, M., Shah, S., Tatlah, I.A.: Impact of principals/directors' leadership styles on job satisfaction of the faculty members: perceptions of the faculty members in a public university of Punjab, Pakistan. *J. Res.* **7**(2), 97–112 (2013)

- Anwar, K., Ghafoor, C.: Knowledge management and organizational performance: a study of private universities in Kurdistan. *Int. J. Soc. Sci. Educ. Stud.* **4**(2), 53 (2017)
- Anwar, K., Louis, R.: Factors affecting students 'anxiety in language learning: a study of private universities in Erbil, Kurdistan. *Int. J. Soc. Sci. Educ. Stud.* **4**(3), 160 (2017)
- Avolio, B.J., Walumbwa, F.O., Weber, T.J.: *Leadership: Current Theories* (2009)
- Bass, B.M., Avolio, B.J.: *Improving Organizational Effectiveness through Transformational Leadership*. Sage, Thousand Oaks, CA (1993)
- Bass, B.M., Stogdill, R.M.: *Handbook of leadership. Theory, Research & Managerial Applications*, vol. 3 (1990)
- Bass, B.M., Avolio, B.J., Jung, D.I., Berson, Y.: Predicting unit performance by assessing transformational and transactional leadership. *J. Appl. Psychol.* **88**(2), 207 (2003)
- Chowdhury, R.: A study on the impact of leadership styles on employee motivation (2014)
- Church, D.M.: Leadership style and organizational growth: A correlational study. Ph.D. thesis, University of Colorado. Fiedler and. E: 1996, „Research on leadership selection and training: One view of the future”. *Administrative Science Quarterly* pp. 241–250. Commitment: An empirical study of selected organizations in corporate sector (Ph.D). Dr D. Y. Patil Educational and psychological measurement, 30(3), 607–610 (2012), from http://www.ashese.co.uk/files/Mengesha_ISSUE_3_cc.pdf
- Dawson (2002). *Practical Research Methods: A user-friendly guide to mastering research techniques and projects*. How To Books Ltd, 3Newtec Place, Press
- AL-Nawafleh, E.A., ALSheikh, G.A.A., Abdulllah, A.A., Bin, A., Tambi, A.M.: Review of the impact of service quality and subjective norms in TAM among telecommunication customers in Jordan. *Int. J. Ethics Syst.* **35**(1), 148–158 (2019)
- Gopal, R., Chowdhury, R.G.: Leadership styles and employee motivation: an empirical investigation in a leading oil company in India. *Int. J. Res. Bus. Manag.* **2**(5), 1–10 (2014)
- Hafeez, M.H., Rizvi, S.M.H., Hasnain, A., Mariam, A.: Relationship of leadership styles, employees commitment and organization performance (a study on customer support representatives). *Eur. J. Econ. Fin. Adm. Sci.* **1**(49), 133–143 (2012)
- Hanifah, H., Susanthi, N.I., Setiawan, A.: The effect of leadership style on motivation to improve the employee performance. *J. Manajemen Transportasi Logistik* **1**(3), 221–226 (2014). <https://doi.org/10.1146/annurev.psych.60.110707.163621>
- Harahsheh, A., Houssien, A., Alshurideh, M., AlMontaser, M.: The effect of transformational leadership on achieving effective decisions in the presence of psychological capital as an intermediate variable in private Jordanian universities in light of the corona pandemic. *Eff. Coronavirus Dis. (COVID-19) Bus. Intell.* **334**, 221–243 (2021)
- Hughes, R.L., Ginnett, R.C., Curphy, G.J.: *Leadership: Enhancing the Lessons of Experience*. McGraw-Hill/Irwin, New York (2015)
- Hussain, T., Ali, W.: Effects of servant leadership on flowers job performance. *Sci., Tech. and Dev.*, 31, 359–368. Investigation in Haramaya University. *Ashese Journal of Business Management*, 1(3), 2, 3. Retrieved (2012).
- Jerotich, T.: Influence of head teachers leadership styles on Employees in secondary school access to their rights In Nandi East Kenya. Ph.D. Thesis (2013)
- Jordanian Telecommunication Sector (2023). <http://www.trc.gov.jo>
- Joseph, E.E., Winston, B.E.: A correlation of servant leadership, leader trust, and organizational trust. *Leadersh. Organ. Dev. J.* **26**(1), 6–22 (2005)
- Krejcie, R.V., Morgan, D.W.: Determining sample size for research activities (1970)
- Kurdi, B., Alshurideh, M., Alnaser, A.: The impact of employee satisfaction on customer satisfaction: theoretical and empirical underpinning. *Manag. Sci. Lett.* **10**(15), 3561–3570 (2020)
- Mengesha, A.: Impact of leadership approaches on employee motivation: an empirical investigation in Haramaya University. *AshEse J. Bus. Manag.* **1**(3), 028–038 (2015)

- Molero, F., Cuadrado, I., Navas, M., Morales, J.F.: Relations and effects of transformational leadership: a comparative analysis with traditional leadership styles. *Spanish J. Psychol.* **10**(2), 358–368 (2007)
- Nuseir, M.T., Al Kurdi, B.H., Alshurideh, M.T., Alzoubi, H.M.: Gender discrimination at work-place: do artificial intelligence (AI) and machine learning (ML) have opinions about it. In: *The International Conference on Artificial Intelligence and Computer Vision*, vol. 1377, pp. 301–316. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-76346-6_28
- Obeidat, U., Obeidat, B., Alrowwad, A., Alshurideh, M., Masadeh, R., Abuhashesh, M.: The effect of intellectual capital on competitive advantage: the mediating role of innovation. *Manag. Sci. Lett.* **11**(4), 1331–1344 (2021)
- Odeh, R.B.M., Obeidat, B.Y., Jaradat, M.O., Alshurideh, M.T.: The transformational leadership role in achieving organizational resilience through adaptive cultures: the case of Dubai service sector. *Int. J. Prod. Performance Manag.* (2021). <https://doi.org/10.1108/IJPPM-02-2021-0093>
- Oreg, S., Berson, Y.: Leadership and Employees 'reactions to change: the role of leaders' personal attributes and transformational leadership style. *Pers. Psychol.* **64**(3), 627–659 (2011)
- Rae, A.: *Organizational behavior: an introduction to your life in organizations*, Pearson International edition, Pearson Research, and Future Directions. *Ann. Rev. Psychol.* **60**(1), 421–449 (2008)
- Telecommunication Regulatory Commission (2022). <http://www.trc.gov.jo>
- Uddin, M.R.: Impact of leadership style on employee motivation: a study on the employee serving in banking organization in Bangladesh. *Int. J. Bus. Market. Manag.* **4**(7), 42–48 (2019)

Author Index

A

Abajaddi, Nesrine 158
Abakarim, Sana 317
Abdalla, Hamed Omar 447
Abdelghafar, Sara 399
Abdelkader, Mohamed F. 487
Abdellah, Ezzati 282
Abdellah, Jamali 553
Abdelsalam, Abdelazeem A. 487
Abdul-Hamid, Yasser 27
Abel, Marie-Hélène 368
Abid, Mohamed 498
Aboulmira, Amina 416
Ahmed, Hanan 119
Ait Rai, Khaoula 407
Al Fouri, Areeg 91
Al Fouri, Shatha 91
Alami Merrouni, Zakariae 107
Alfour, Areeg 459
Ali, Abdulrahman 399
Al-lawam, Haron Ismail 594
Al-lawama, Haron Ismail 447
Al-Maamari, Asaad Ali Muslim 379
Almahairah, Mohammad Salameh Zaid 447, 594
AlMahdawi, Asmaa Jumaha 447
Almomani, Hiba Hussein Mohammad 594
Al-Shaar, Anwar Saud 447, 594
Alshurideh, Muhammad Turki 91, 379, 447, 594
Alshurideh, Muhammad 459
Alwaely, Suad Abdalkareem 379
Amarjouf, Madiha 68
Amine, Baina 179
Amraouy, Mohammed 327
Amzil, Abdellah 498

Antari, Jilali 407
Aouad, Siham 575
Aravinda, C. V. 294
Atya, Hanan 58

B

Badr, Nagwa L. 16
Bahja, Fadoua 68
Baina, Amine 138
Bellafkih, Mostafa 138, 327
Bellar, Oumaima 138
Bennane, Abdellah 327
Beshir, Sara 337
Bhatnagar, Roheet 294
Bhattacharyya, Mousumi 358
Bhattacharyya, Siddhartha 524
Bhuvaneswari, S. 231
Bouchra, Frikh 77
Bousarhane, Btissam 37
Bouzidi, Driss 37
Bouzidi, Driss 48
Brahim, Ouhbi 77

C

Chami, Mouhcine 68

D

Dang, Doan Thai 149
Darawsheh, Saddam Rateb 447, 594
Darif, Anouar 129, 242
Darwish, Ashraf 399
Debnath, Narayan C. 149, 188
Di-Martino, Joseph 68
Dutta, Pushan Kumar 305
Dutta, Tulika 524

E

- Ebied, Hala M. 272
 El Kafhali, Said 197, 477, 498
 El Khadiri, Zakaria 262
 Elfahm, Youssef 158

F

- Farchi, Abdelmajid 158
 Fayed, Ahmed Rabie 508
 Frikh, Bouchra 107

G

- Gamal, Gehan 487
 Ganesan, Subramaniam 436
 Ghaleb, Moshira S. 272
 Ghosh, Anupam 358
 Gmach, Imen 447
 Gouspillou, Philippe 368
 Guirguis, Shawkat K. 216

H

- Hafdi, Zakaria Soufiane 197
 Hamad, Safwat 119
 Hamed, Esraa A.-R. 16
 Hamour, Randa Abu 459
 Hanini, Mohamed 498
 Hassouni, Larbi 565
 Hrimech, Hamid 416
 Hussein, Ashraf S. 119

I

- Ibn-Elhaj, El Hassan 68

K

- Khalifa, Nour Eldeen M. 508
 Khaoula, Elhabyb 179
 Kim, Tai-hoon 535
 Kotb, Amira 508
 Krishnaveni, M. 164
 Kumar, Gulshan 535

L

- Lachgar, Maryem 129, 242
 Lachgar, Mohamed 416
 Larhlimi, Ibtissam 129, 242

- Latif, Rachid 262
 Le, Ngoc Luyen 368

M

- Machkour, Mustapha 407
 Madbouly, Magda M. 216
 Magdy, Ahmed 58, 206, 487
 Manal, Hilali 282
 Martis, Jason Elroy 294
 Meriem, Hnida 348
 Midya, Sadip 358
 Mikram, Hind 477
 Mitra, Anirban 358
 Moawad, Mayar M. 216
 Mohamed, Radwa 206
 Mohammed, Ammar 27
 Mohammed, Heba 58
 Mohana, M. 3
 Mostafa, Bellaflkih 179
 Mostafa, Lamiaa 337
 Moujahid, Fatima Ezzahra 575
 Mouncif, Hicham 129, 242
 Mounir, Badia 158

N

- Nafea, Sherif F. 206
 Naji, Zakaria 48
 Najib, Naja 553
 Najima, Daoudi 348
 Narmadha, V. 164
 Ngo, Ngoc Phuong 188
 Nguyen, Thanh Phuong 251
 Nguyen, Vinh Dinh 149, 188
 Nikolov, Nikola S. 426

O

- O'Mahony, Laura 426
 O'Sullivan, David JP 426
 Ouchitachen, Hicham 129, 242
 Ouf, Mahmoud 27
 Ouhibi, Brahim 107

P

- Panigrahi, Bijaya Ketan 524
 Pham, Van Tien 251

Q

Qassimi, Sara 317

R

Raafat, Kareem 58

Raajeswari, Paa. 164

Rabbah, Jalal 565

Rachida, Ajhoun 348

Rakrak, Said 317

Ramya, P. 164

Rather, Irfan Ahmad 535

Ravi, Renjith V. 305

Ridouani, Mohammed 565

Rima, Sandoussi 348

Roy, Asmita 358

Roy, Sudipta 305, 358

S

Saadi, Youssef 477

Saddik, Amine 262

Saha, Rahul 535

Said, Ben Alla 282

Sakher, Shatha 91

Salem, Mohammed A.-M. 16

Sameh, Ahmed 58

Sannidhan, M. S. 294

Shaaban, Sara 58

Shajrawi, Ahmad A. I. 594

Shater, Azhar 447

Shedeed, Howida A. 119

Subashini, P. 3, 164, 231

T

Taha, M. H. N. 508

Talaghzi, Jallal 327

Tolba, Mohamed F. 16, 272

Tran, Thanh Hoang 149

Tyagi, Ishan 436

Y

Yassine, Sembati 553

Youssef, Bekach 77

Z

Zbakh, Mostapha 575