Check for updates

Unlocking the potential of RNN and CNN models for accurate rehabilitation exercise classification on multi-datasets

Moamen Zaher[1,5] · **Amr S. Ghoneim[2]** · **Laila Abdelhamid[3]** · **Ayman Atia[4,5]**

Abstract

Physical rehabilitation is crucial in healthcare, facilitating recovery from injuries or illnesses and improving overall health. However, a notable global challenge stems from the shortage of professional physiotherapists, particularly acute in some developing countries, where the ratio can be as low as one physiotherapist per 100,000 individuals. To address these challenges and elevate patient care, the field of physical rehabilitation is progressively integrating Computer Vision and Human Activity Recognition (HAR) techniques. Numerous research efforts aim to explore methodologies that assist in rehabilitation exercises and evaluate patient movements, which is crucial as incorrect exercises can potentially worsen conditions. This study investigates applying various deep-learning models for classifying exercises using the benchmark KIMORE and UI-PRMD datasets. Employing Bi-LSTM, LSTM, CNN, and CNN-LSTM, alongside a Random Search for architectural design and Hyper-parameter tuning, our investigation reveals the (CNN) model as the top performer. After applying cross-validation, the technique achieves remarkable mean testing accuracy rates of 93.08% on the KIMORE dataset and 99.7% on the UI-PRMD dataset. This marks a slight improvement of 0.75% and 0.1%, respectively, compared to previous techniques. In addition, expanding beyond exercise classification, this study explores the KIMORE dataset's utility for disease identification, where the (CNN) model consistently demonstrates an outstanding accuracy of 89.87%, indicating its promising role in both exercises and disease identification within the context of physical rehabilitation.

Keywords Physical rehabilitation exercises · Deep learning · Hyper-parameter tuning · LSTM · Bi-LSTM · CNN · CNN-LSTM

1 Introduction

Physical rehabilitation exercises are crucial for individuals to recover body function and capabilities following an injury, illness, or surgery. The exercises are critical components of recovery programs intended to help patients regain their lost physical abilities, such as

✉ Moamen Zaher
moamenibrahim_psw@fci.helwan.edu.eg; moibrahem@msa.edu.eg

Extended author information available on the last page of the article

🖄 Springer

flexibility and strength, and to facilitate a safe return to their normal daily activities. This research explores how computer vision and the latest deep-learning techniques can be used to evaluate physical rehabilitation exercises. The findings will help promote healing, prevent future injuries, and improve overall health and well-being.

The importance of physical therapy research is emphasized by global health trends, according to the World Health Organization (WHO) [1]. Currently, an astounding estimate of 2.4 billion individuals worldwide require rehabilitation for their health conditions. Notably, over 50% of individuals who need essential rehabilitation services lack access to them. This severe lack of accessibility is worsened by insufficient funding and undervaluation of rehabilitation services, particularly in countries with inadequate healthcare systems. Low- and middle-income countries have less than ten skilled rehabilitation practitioners per 1 million people, which is critically insufficient. Furthermore, with the projected doubling of the global population by 2050, there is an urgent need to address the increasing demand for rehabilitation services, particularly in developing countries where the strain on healthcare systems is becoming more pronounced. This contextual backdrop highlights the importance of advancing physical therapy research to fill gaps and improve the effectiveness of rehabilitation services globally.

The integration of digital technologies and Artificial Intelligence (AI) into home-based rehabilitation, particularly via exercise classification, showcases its groundbreaking potential. The application of auxiliary systems and AI is critical for tailoring exercises and monitoring progress in shoulder rehabilitation treatments [2, 3]. The incorporation of physical activity into rehabilitation routines, as Loellgen et al. [4] proposed, emphasizes the critical role of structured exercises in improving recovery outcomes.

Furthermore, innovative strategies like motor imagery and action observation [5], coupled with the introduction of digital tools in orthopedic recovery [6], and the adoption of virtual reality and digital biofeedback systems for post-operative rehabilitation [7, 8], collectively signal a move towards more personalized and accessible rehabilitation options. These advancements not only confirm the utility of exercise classification in home-based rehabilitation systems but also underscore a transition towards solutions that prioritize patient-centric care, efficiency, and accessibility. Additionally, these technological advancements align with the Sustainable Development Goals (SDGs), particularly SDG 3, which seeks to ensure health and well-being for all at every age. This approach not only improves patient outcomes but also aligns with the World Health Organization's (WHO) goals for universal health coverage by enhancing the accessibility and efficiency of rehabilitation services.

Recently, there has been a growing emphasis on Human Activity Recognition (HAR) research, particularly within physical rehabilitation. For instance, in [9], the authors present a comprehensive review of Indoor-HAR, an emerging technology leveraging cameras and vision-based sensors to assess human actions within indoor settings. The review encompasses contemporary approaches, existing challenges, practical implementations, dataset resources, and a proposed taxonomy for Indoor-HAR. The study also highlights the potential of this technology in various industries, such as healthcare, surveillance, and human-computer interaction. It has multiple applications in smart homes, elderly care, and assisted living. Debnath et al., [10] investigated different computer vision approaches for physical rehabilitation and assessment. This has facilitated the creation of novel procedures that aid in the monitoring and evaluation of physical rehabilitation exercises, providing benefits to patients receiving rehabilitation for injuries or disabilities. Recent advancements in deep learning algorithms have enhanced the accuracy and efficiency of HAR systems. The current and prospective applications of Artificial Intelligence (AI) in physiotherapy and rehabilitation for advancing medical digitization are explored in [11]. On the other hand, [12] examines the application of

a subset of machine learning approaches, namely, Reinforcement Learning. Virtual Reality (VR) approaches are also applied in this domain for physiotherapists and physicians [13]. VR approaches construct an intelligent space using multi-camera systems to capture gait data, thus creating a 3D virtual environment. Several studies in this discipline are skeleton-based [14] or pose-guided [15].

This study focuses on Pose-guided Activity Recognition, which involves the analysis of human body poses to understand and classify various activities, uncovering their potential benefits for patients and healthcare providers. In the accompanying image as shown in Fig. 1 where circles depict body joints, and lines represent intra-body edges, capturing the natural connections in the human body. This visual representation, reminiscent of Kinect pose frames over time, illustrates the temporal sequence of human poses. The duplication of these poses in each frame allows for recognizing and analyzing activities based on the moving body positions. This approach is a crucial aspect of the broader field of human activity recognition.

Popular deep learning algorithms have been utilized for Human Activity Recognition (HAR). Numerous researches utilized Long Short-Term Memory (LSTM) for Human Activity Recognition [16–19]. Bidirectional LSTM (BiLSTM) has been used before for Human Activity Recognition and produced solid performance and acceptable results [20–22]. Significant contributions to the field of Human Activity Recognition leveraging Convolutional Neural Network (CNN) architectures can be found in the research of [23–25], where the utilization of (CNN) models exhibited remarkable efficacy in the precise identification and classification of human activities. Wang et al., [26], and Yadav et al., [27], have applied (CNN-LSTM) for Human Activity Recognition and proved its superiority. Various deep learning

Fig. 1 Skeleton-based activity recognition demonstrates circles representing body joints and lines defining intra-body edges based on natural connections in human bodies. The duplication of human poses replicates the time sequences across frames captured over time
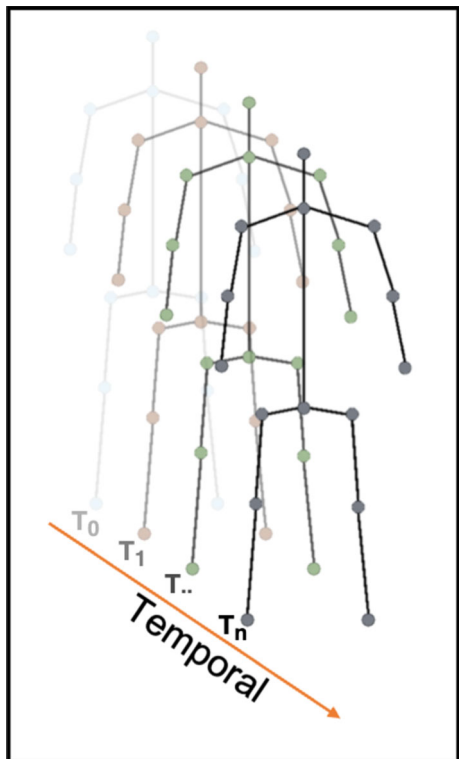
techniques, such as transfer learning, were utilized. Transfer Learning is a machine learning technique where knowledge or insights acquired from one dataset or task are applied to another closely related dataset or task. This technique entails fine-tuning a pre-trained model on a new dataset or task, frequently used when the new task or dataset has limited data or computational resources, making it difficult to train a new model from scratch. Numerous researchers have applied transfer learning in human activity recognition, particularly rehabilitation exercises [14, 15, 28].

Additionally, transform learning, a feature extraction or dimensionality reduction technique that involves converting high-dimensional data into a lower-dimensional space, has been used in this domain. Transformers models are useful for decreasing the computational complexity of working with high-dimensional data and extracting significant features that can be used for a specific job. For example, Réby et al. applied Transform Learning in this domain using the Graph Transformer approach [29].

Real-time healthcare applications pose a significant challenge due to their demand for substantial computational resources, a concern that has garnered increased attention in recent healthcare application research. Various studies have endeavored to address this challenge by achieving a 55% reduction in delay, a 72% improvement in response time, and a 70% reduction in energy consumption [30].

Our proposed model offers the potential for seamless deployment within a mobile application, empowering patients to engage in rehabilitative exercises within the comfort of their homes, utilizing a Kinect camera to capture and feed relevant data to the model. The models determine the accuracy of the exercise execution, a pivotal aspect for expediting the recovery process. This at-home exercise monitoring circumvents the necessity for physical visits to physiotherapy clinics. This paradigm has gained heightened significance during the ongoing global pandemic, where restrictions on outdoor activities are prevalent. This prioritizes patient safety and alleviates the strain on physiotherapists, a profession already grappling with a global shortage of practitioners.

Video-based techniques for feature extraction are resource-intensive [31–33], while manual feature extraction from videos for 2D image conversion is less demanding but often less robust [34]. Our methodology modifies the 1D approach from [35], integrating a novel feature extraction method that balances performance with resource efficiency. This hybrid strategy is particularly cost-effective for at-home rehabilitation, enabling deployment on low-power devices like mobile apps for real-time patient feedback.

The main contributions of this paper are :

1. **Feature Engineering:** The study introduces a novel approach to represent exercises as 1D vectors through comprehensive feature engineering, employing various statistical techniques.
2. **Comparative Study:** A detailed comparative analysis is conducted, assessing the performance of four distinct models (LSTM, Bi-LSTM, CNN, and CNN-LSTM) across two diverse datasets. This comparative study provides insights into the strengths and weaknesses of each model in the context of exercise classification.
3. **Disease Classification:** The models developed in the study are further utilized for disease classification based on five specific exercises from the KIMORE dataset. This application demonstrates the proposed models' versatility and potential clinical relevance beyond exercise classification.

This manuscript is structured into eight sections: Introduction, Related Work, Dataset, Methodology, Experiments, Results, Discussion, Conclusion and Future Work.

2 Literature review

This section is organized into four subsections: Deep Learning-Based, Ensemble-Based, Transformer-Based, and Transfer-Based Approaches. Following the approach taken by Akkem et al., [36], a comparative analysis of specific methods within these approaches has been conducted. Table 1 provides a comprehensive overview of various models applied in HAR along with their associated evaluation metrics, results, datasets, and identified limitations. One research gap evident from the table is the common challenge of real-time applicability, particularly in resource-constrained environments. Several models, such as Graph Convolutional Siamese Network [37], Ensemble CNN-RNN [38], Ensemble-based Graph Convolutional Network [39], and Graph Attention Network [40], exhibit high computational costs, potentially limiting their suitability for real-time deployment on devices with constrained resources. Additionally, some models, including CNN-LSTM [34], LSTM-1D CNN [35], 3D CNN [41], and VGG-16 [28], highlight the need for further optimization to improve accuracy, indicating an avenue for future research in model refinement.

The state-of-the-art models in human activity recognition and physical rehabilitation, such as CNN-LSTM, Graph Convolutional Siamese Networks, LSTM-1D CNN, 3D CNN, Ensemble CNN-RNN, Ensemble-based Graph Convolutional Network, Graph Attention Network, and VGG-16, demonstrate significant advancements in accuracy and performance across various datasets. However, they also reveal critical limitations concerning generalizability, computational efficiency, and scalability. These challenges highlight the ongoing need for optimization to balance model complexity with practical applicability, ensuring that advanced deep learning models can be deployed effectively in real-world settings without being hindered by computational constraints or loss of relevant information due to preprocessing choices.

Another noteworthy research gap involves transforming time-series data into 2D images or RGB formats, as Boukhennoufa et al., [28] proposed. This process introduces feature engineering choices that may encode biases or lead to information loss, impacting the model's generalizability. Additionally, it is observed that several studies lack experimentation on multiple datasets, limiting the generalizability of their findings. Addressing these gaps can contribute to developing more efficient and accurate HAR models with improved real-time capabilities and enhanced generalizability.

Our methodology aims to mitigate the constraints outlined in the preceding Table 1. The refinement of model architectures was executed through a random search process, prioritizing attaining accuracy to ensure broad generalization. Extensive testing across multiple datasets incorporated a robust 5-fold cross-validation strategy to validate the models' performance. Leveraging deep learning models was instrumental in addressing the inference time challenges associated with attention-based and ensemble models. We opted for a 1D approach over 2D, strategically sidestepping the complexities associated with image representation choices and the intricate nature of 2D models.

2.1 Deep learning-based approaches

Kumar et al., [38] introduced an innovative Deep-HAR model, Fusing Convolutional Neural Networks (CNNs) to extract features and Recurrent Neural Networks (RNNs) to capture temporal patterns within sequential data. Their evaluation encompassed three publicly accessible datasets (WISDM [44], PAMAP2 [45], and KU-HAR [46]), revealing exceptional performance across all activity types. Notably, the model achieved remarkable accuracy rates,

Table 1 Limitations and state-of-the-art status

Ref.	Model	Metric	Result	Dataset	Limitation
[34]	CNN-LSTM	Average Absolute Deviation	0.02527	UI-PRMD[42]	Generalizability & Consumes more time than (CNN)
[37]	Graph Convolutional Siamese Network	Accuracy	99.2%, 97.0%	UI-PRMD, IntelliRehabDS	The GCN and Siamese network architecture, while effective, can be computationally expensive. This might limit its real-time applicability on resource-constrained devices in clinical settings.
[35]	LSTM-1D CNN	Accuracy	92.33%	KIMORE[43], UI-PRMD, IntelliRehabDS	The model's results suggest further optimization is necessary to improve accuracy.
[41]	3D CNN	Accuracy	90.57%	KIMORE	The model's results suggest further optimization is necessary to improve accuracy.
[38]	Ensemble CNN-RNN	Accuracy	99.98%, 99.64%, 99.98%	WISDM[44], PAMAP2[45], KU-HAR[46]	Complex and computationally expensive, potentially limiting its real-time applicability on resource-constrained devices
[39]	Ensemble-based Graph Convolutional Network	Accuracy	86.9%, 80.1%	UI-PRMD, KIMORE	Complex and computationally expensive, potentially limiting its real-time applicability on resource-constrained devices

Table 1 continued

Ref.	Model	Metric	Result	Dataset	Limitation
[40]	Graph Attention Network	Accuracy	96.4%, 88.2%, 61.0%	NTU60, NTU120, Kinetics-400	High computational cost, due to self-attention mechanisms and dynamic graph construction, hinders its real-time
[28]	VGG-16	Accuracy	98.53%	SBHAR [47]	Transforming the time-series data into 2D images or RGB formats involves feature engineering choices. These choices could potentially encode biases or lose relevant information from the original data, affecting the model's generalizability

scoring 99.98% for simple activities, 99.64% for complex activities, and 99.98% for heterogeneous activities. Similarly, our study employs various (CNN) and RNN models to address a specific healthcare concern, focusing on rehabilitation, and evaluates their efficacy on two dedicated rehabilitation datasets.

Tasmin et al., [48] focused on complex and diverse human activities rather than simple activities like walking or sitting. Using skeletal data, they introduced a Dynamic Edge Convolutional Neural Network (DECNN). DECNN operates on a graph representation of skeletons, treating joints as nodes and modeling their connections as edges. The authors evaluated their approach on UTD-MHAD [49] and MSR-Action3D [50], achieving high accuracy.

The integration of a CNN-LSTM hybrid model is a prevalent approach in both human activity recognition [26, 27, 34, 51–53], and disease classification [54]. Lakhan et al. employed a CNN-LSTM architecture in proposing a framework for detecting Autism Spectrum Disorder. Liao et al., [34] also introduced a deep learning architecture that combines CNN and LSTM networks to evaluate physical rehabilitation exercises. Their framework proficiently classifies exercises in the UI-PRMD dataset, demonstrating practicality with low absolute deviation per exercise.

2.2 Ensemble-based approaches

Ensemble learning, as investigated by Zhang et al. [55] extensively, combines data fusion, modeling, and mining into a unified framework. Effective ensemble methods carefully combine members to enhance performance, avoiding random fusion issues. These methods are categorized (in classification tasks) into data-level, feature-level, decision-level, and model-level approaches. Chenguang et al., [56] utilize ensemble learning for hand function assessment, while Chihiro et al., [57] apply it for predicting functional outcomes after spinal cord injury. Additionally, Wenchuan et al., [58] leverage ensemble learning for personalized remote training in Parkinson's disease patients. In the realm of skeleton-based rehabilitation exercises, EGCN, an ensemble-based framework by Yu et al., [39], demonstrates robust performance on both UI-PRMD and KIMORE datasets.

2.3 Transformer-based approaches

In this paper [59], the authors present a new Spatial-Temporal Graph Convolutional Networks (ST-GCN) model for dynamic skeletons. ST-GCN is designed to autonomously learn spatial and temporal patterns from skeleton data to enhance human action recognition. The authors employ graph convolutional networks (GCNs) with spatial and temporal convolution operations to achieve this. These operations capture the dynamic aspects of human body skeletons, represented as spatial and temporal graphs. These graphs are constructed based on the inherent connections between human body parts and the temporal progression of skeleton sequences. Furthermore, the authors introduce a partitioning strategy to segment the spatial graph into distinct components, aligning with different body parts' physical significance and functions. The effectiveness of the proposed approach is extensively evaluated on two large datasets, Kinetics and NTU-RGBD, resulting in state-of-the-art performance. Notably, the method achieves an accuracy of 81.5% on NTU-RGBD (CV) and 30.7% on Kinetics-Skeleton. The authors conclude that this approach can be readily extended to other tasks involving skeleton data, including gesture recognition, pose estimation, and human-object interaction.

Rahevar et al. proposed a Spatial-Temporal Dynamic Graph Attention Network (ST-DGAT) [40], which presents a groundbreaking approach to extracting valuable features from human body skeleton data. In contrast to previous methods like Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT), which struggled with capturing long-range joint dependencies, (ST-DGAT) introduces dynamic graph attention. By reordering the weighted vector operations in (GAT), (ST-DGAT) significantly enhances the attention mechanism, allowing it to learn spatial-temporal patterns from skeleton sequences effectively. The model outperforms existing approaches and achieves state-of-the-art accuracy on large-scale datasets, including NTU60, NTU120, and Kinetics-400, with accuracy levels of 96.4%, 88.2%, and 61.0%, respectively.

Reby et al., [29] proposed a graph transformer model for assessing the effectiveness of physical rehabilitation on the UI-PRMD dataset [42]. Their model combines self-attention mechanisms with spatiotemporal graph networks, achieving state-of-the-art binary classification and quality score prediction results. It is essential to emphasize that these outcomes are specific to a singular dataset, and there is a possibility that the performance of the (G2PRE) model may differ on other datasets.

Peng et al., [14] proposed a personalized rehabilitation training recognition framework using transfer learning based on human skeleton data. It employs pre-trained (CNN) models for feature extraction and hierarchical classification for exercise recognition. The approach achieved significant accuracy improvements for personalized rehabilitation training.

Generally, Transformers are capable of modeling long-range dependencies in sequential data. However, it requires significant computational costs due to self-attention mechanisms. This can hinder their real-time applicability on resource-constrained devices often used for (HAR). (CNNs), in contrast, offer efficient spatial feature extraction, making them well-suited for HAR tasks with fixed-size sensor inputs. However, their limitations in handling long temporal dependencies might necessitate careful design and segmentation strategies. (RNNs), specifically (LSTMs), provide a balance between sequential modeling and computational efficiency, but their performance can still be impacted by sequence length and model complexity.

2.4 Transfer-based approaches

Boukhennoufa et al., [28] introduced an innovative strategy to enhance activity recognition accuracy in post-stroke rehabilitation assessments, incorporating the Geometric Mean of Absolute and Relative Frequencies (GMAF) [60]. The proposed methodology encompasses two primary phases: feature extraction and classification. GMAF is applied to extract features from raw accelerometer data in the feature extraction stage. Subsequently, the classification phase involves experimenting with three different approaches: 1D CNN, 2D CNN, and a transfer learning approach using VGG-16 [61]. Implementing a basic 1D (CNN) classifier on these images elevates the accuracy of test data from 94% using traditional segmentation to 97.06%. Additionally, the transformation of 2D images into RGB format, coupled with using a 2D (CNN) classifier, results in a noteworthy accuracy increase to 97.52%.

However, certain limitations accompany this approach. It demands additional prepossessing to convert data into images. Although the 1D (CNN) achieves satisfactory results, it significantly requires fewer computational resources than VGG models. Using transfer learning with pre-trained models, such as (VGG16), designed for different tasks, like image recognition, may not fully capture the specific features and dynamics of post-stroke activity data. This potential domain mismatch could limit the benefits derived from transfer learning.

In summary, the choice of deep learning architecture for HAR necessitates a delicate balance between performance and computational constraints. Transformers, adept at capturing long-range dependencies in sequential data, come with significant computational costs attributed to self-attention mechanisms. Requiring significant computational can impede their real-time applicability on resource-constrained devices commonly utilized for HAR. On the other hand, CNNs excel in efficient spatial feature extraction, rendering them suitable for HAR tasks with fixed-size sensor inputs. However, their limitations in handling long temporal dependencies may require meticulous design and segmentation strategies. RNNs, particularly LSTMs, balance sequential modeling and computational efficiency, but their performance may still be influenced by sequence length and model complexity. Ensemble models, while offering a potential solution to prior challenges, come at the expense of increased inference time as at least two models are run for the same classification task. Transfer Learning, although promising, can face obstacles due to domain dissimilarity, necessitating task-specific fine-tuning. In conclusion, the optimal HAR architecture selection demands a comprehensive understanding of the trade-offs between accuracy, computational resources, and the temporal characteristics of sensor data.

3 Datasets characteristics

This study employs two separate datasets to validate our action recognition methodology. The first dataset, known as the UI-PRMD dataset [42], and the second dataset, referred to as the KIMORE dataset [43].

3.1 UI-PRMD

The UI-PRMD dataset, introduced by Vakanski et al. in their research [42], represents a comprehensive resource dedicated to a physical rehabilitation experiment. Ten healthy individuals participated, performing ten rehabilitation exercises with ten repetitions, each with correct and wrong techniques for each exercise, resulting in 20 classes and 2000 records. Two motion-capturing systems, a Vicon optical tracker and a Kinect camera, were utilized throughout the study. Including both techniques enhances the dataset's utility for training and evaluating models tailored to physical rehabilitation. Providing insights into the correct form and identifying common errors, the UI-PRMD dataset is crucial in advancing research in human activity recognition and rehabilitation analysis. Within this paper, the classification of correct and erroneous techniques holds paramount importance, aligning with the primary objective of facilitating accurate assessments during at-home exercises.

The investigation specifically targeted ten distinct exercises: Deep Squat, Side Lunge, Standing Shoulder Abduction, Hurdle Step, Standing Shoulder Scaption, Standing Shoulder Extension, Inline Lunge, Sit Stand, Standing Active Straight Leg Raise, and Standing Shoulder Internal-External Rotation. Notably, each exercise was executed with correct and incorrect techniques, providing a nuanced dataset for analysis. The complete listing of the ten exercise classes can be found in Table 2.

The UI-PRMD dataset records data in the YXZ coordinate system, where Y denotes height, X represents width, and Z indicates depth. Various body parts, including Chest, Waist, Head, Left Collar, Right Collar, Left Upper Arm, Right Upper Arm, Left Forearm, Right Forearm, Head tip, Neck, Left Upper Leg, Right Upper Leg, Left Hand, Right Hand, Left Foot, Right

Table 2 Compilation of exercises and corresponding labels in the UI-PRMD Dataset

Label	Exercise Name	Label	Exercise Name
Ex 1	Deep Squat	Ex 6	Standing Leg Raise
Ex 2	Hurdle Step	Ex 7	Shoulder Abduction
Ex 3	Inline Lunge	Ex 8	Shoulder Extension
Ex 4	Side Lunge	Ex 9	Shoulder Internal-External Rotation
Ex 5	Sit to Stand	Ex 10	Shoulder Scaption

Foot, Spine, Left Leg Toes, and Right Leg Toes, were utilized to extract joint information and corresponding angles from the sensor data.

3.2 KIMORE

Marianna Capecci et al., [43] conducted a study titled "Kinematic Analysis of Movements and Clinical Assessments for Remote Supervision of Physical Rehabilitation (KIMORE)" at the University of Pisa, Italy. Employing RGB-D sensors, specifically the Kinect v2, the study recorded RGB and depth videos and skeletal joint positions during five targeted exercises addressing lower back pain. Data collection occurred in a controlled environment using Kinect cameras, incorporating a diverse set of participants, including both healthy individuals and patients with various motor dysfunctions. This diversity is critical for developing algorithms with robust generalization capabilities across different clinical populations. The dataset comprises 78 participants, including 44 individuals without identified health issues, consisting of 17 Expert Professionals and 27 Not Experts. Additionally, there are 34 participants with motor dysfunctions, representing common neurological and musculoskeletal conditions such as stroke, Parkinson's disease, or low back pain. This diverse representation ensures the dataset's applicability to various rehabilitation settings.

The primary purpose of the dataset is exercise classification, with an additional experiment demonstrating the feasibility of disease classification. The five exercises in the dataset are listed in Table 3.

Similarly to Reby et al., [29], we prioritize skeleton-based methods because of their resilience to variations in body scales, motion speeds, camera perspectives, and environmental interference compared to RGB image-based methods. In line with this preference. In this study, we primarily employ the Kinect-extracted data, which performs better than directly extracting data from individual joints, as evidenced by previous research [35]. Furthermore, the dataset includes a set of meticulously defined features for each exercise, which medical experts established to characterize the scope of each exercise. These features, validated against

Table 3 Compilation of exercises and corresponding labels in the KIMORE Dataset

Label	Exercise Name
Ex 1	Lifting of arms
Ex 2	Lateral tilt of the trunk with the arms in extension
Ex 3	Trunk rotation
Ex 4	Pelvis rotations on the transverse plane
Ex 5	Squatting

a stereophotogrammetric system, can be analyzed to compute a performance score for each participant. The dataset is structured into three sub-folders: Raw data, Script, and Label. The participant cohort is further categorized into two main macro-groups: the Pain and Posture disorders Group (GPP) and the Control Group (CG). Within the CG, there is a subdivision into two subgroups, namely CG with expertise in physiotherapy exercises (CG-E) and CG without such expertise (CG-NE). Similarly, the GPP is divided into three sub-groups based on specific diagnoses, namely Parkinson's disease, Stroke, and Low Back Pain. The dataset was employed for exercise classification, and an additional experiment was undertaken to showcase the feasibility of disease classification.

4 Methodology

Figure 2 illustrates a system overview, capturing the entire workflow from data acquisition to deployment. The diagram provides a detailed depiction of key stages, including preprocessing, Hyper-parameter tuning, model training, diverse metric-based evaluation, visualization of results, and model deployment. Each of these stages will be intricately expounded upon in their corresponding sections. Section 4.1 delves into the intricacies of data acquisition and alternative sensors, while section 4.2 illuminates the data preparation and preprocessing phase. In section 4.3, we delve into the details of adjusting Hyper-parameter tuning, and the employed classification models are detailed in section 4.4. Section 4.5 explains the different deployment choices that have been considered.

4.1 Data input

The patient initiates the exercise in front of the sensor device, whether a Kinect or a conventional RGB camera. Our investigation primarily focuses on skeletal data acquired through Kinect camera technology; however, our model demonstrates versatility by not relying exclusively on Kinect. An RGB camera can also be employed because our models are constructed utilizing joint coordinates. Nevertheless, additional processing is essential to extract skeletal joints. Two viable approaches for using RGB cameras as a primary sensor are OpenPose [62]
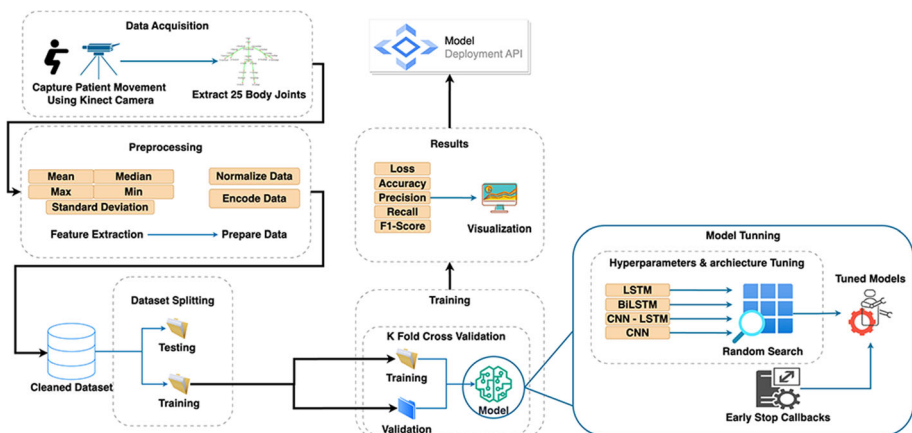

Fig. 2 Overview of the Methodology

and BlazePose [63]. OpenPose and BlazePose are widely adopted solutions for extracting skeletal joints from the human body. BlazePose offers a significantly faster runtime than OpenPose[64], making it suitable for diverse applications such as movement pre-screening and activity classification. BlazePose, functioning as a lightweight and efficient (CNN) model, excels in real-time pose estimation by predicting the 3D pose of an individual from a single image or video frame. The model employs CNNs to extract features from input images, followed by regression layers to predict the location of body key points. On the other hand, MediaPipe [65], a framework based on the BlazePose model, is better suited for constructing real-time machine-learning pipelines.

Despite both MediaPipe and Kinect providing joint coordinates in three dimensions (XYZ), there exists a disparity in the number of joints. MediaPipe extracts 33 joints, while the Kinect's joint count varies depending on the version. For instance, the UI-PRMD dataset utilizes a Kinect, resulting in 22 joints, whereas KIMORE employs a Kinect-v2, yielding 25 joints. Consequently, slight modifications in the input shape are necessary depending on the specific sensor utilized. Subsequently, these joint data are forwarded to the next phase for preprocessing. Figure 2 visualizes the data acquisition phase.

4.2 Data preparation & preprocessing

Data processing played a vital role in this phase. Each dataset's unique structure necessitated different preparatory procedures due to variations in format, features (especially the number of joints), and data volume between the two sets. Nevertheless, uniform preprocessing techniques were applied to ensure unbiased comparisons across all algorithms. The pre-processing phase is visualized in Fig. 2.

In the case of the UI-PRMD dataset, the Kinect Camera records the data of 22 body joints, storing this information in a vector denoted as V. At each time instance t, the representation of each joint data J_n consists of three-dimensional coordinates: X_t, Y_t, and Z_t. Feature extraction techniques are then applied to process each joint data. These techniques include mean, median, minimum, maximum, and standard deviation. Applying these methods results in extracting 330 features (22 body joints x 3 Coordinates x 5 statistical techniques for each joint coordinate).

For the KIMORE dataset, the original dataset comprises XYZ coordinates for 25 distinct body joints, accompanied by a confidence score. This score quantifies the degree of confidence in the 3D joint position as recorded by the Kinect sensor, with values ranging from 1 (indicating a high level of confidence) to 0 (indicating a low level of confidence). Consequently, the dataset incorporates a total of 100 features. Salwa et al., [66] surveyed the most common feature extraction techniques used in human activity recognition. Instead of using only one method for feature extraction, various statistical techniques were employed, encompassing measures such as mean, median, minimum, maximum, and standard deviation, applied to each of the 100 features. This process resulted in an expanded total of 500 features [25 body joints x 4 (3 Coordinates + 1 confidence score) x 5 statistical techniques for each joint coordinate]. It is noteworthy that, unlike the UI-PRMD dataset, it does not exhibit a balanced class distribution. Nevertheless, the disparities between class distributions are marginal, making implementing oversampling or down-sampling techniques unnecessary in the context of exercise classification experiments. The feature vector (V) is formulated in Equation 1; here, the variable n represents the number of body joints, with a value of 25 for the KIMORE dataset and 22 for the UI-PRMD dataset. In this context, i denotes the joint index, y corresponds to vertical, x signifies horizontal, and Z represents depth. It's worth

noting that, for the KIMORE dataset, an additional value was incorporated as a confidence score.

$$
\begin{aligned}
\text{Feature Vector } (V) = &\sum_{i=1}^{n} \left(\text{median}(j_{i_x}) + \text{median}(j_{i_y}) + \text{median}(j_{i_z}) \right) \\
&+ \sum_{i=1}^{n} \left(\text{mean}(j_{i_x}) + \text{mean}(j_{i_y}) + \text{mean}(j_{i_z}) \right) \\
&+ \sum_{i=1}^{n} \left(\text{max}(j_{i_x}) + \text{max}(j_{i_y}) + \text{max}(j_{i_z}) \right) \\
&+ \sum_{i=1}^{n} \left(\text{min}(j_{i_x}) + \text{min}(j_{i_y}) + \text{min}(j_{i_z}) \right) \\
&+ \sum_{i=1}^{n} \left(\text{std}(j_{i_x}) + \text{std}(j_{i_y}) + \text{std}(j_{i_z}) \right)
\end{aligned}
\tag{1}
$$

For the disease classification experiment (the experiment's objectives and constraints are discussed in Section 5.2), we employed diseases as class labels instead of utilizing the five exercises as class labels. We combined the 'Expert' and 'Not Expert' classes into one class, 'Normal,' representing healthy individuals. The remaining classes include Parkinson's disease, Stroke, and Back Pain. It's worth noting that the dataset (for disease classification) exhibited a significant class imbalance, with notable variations in sample sizes among different classes. To address this imbalance and enhance the data distribution, we employed the Synthetic Minority Over-sampling technique (SMOTE) [67]. SMOTE is a machine learning data augmentation method specifically designed to mitigate challenges associated with class imbalance.

The core principle of SMOTE involves generating synthetic instances within the minority class by interpolating between existing data points. This augmentation technique effectively counteracts skewed class distributions, improving model performance, especially in scenarios characterized by substantial class imbalances. A visual representation of the class distribution before and after applying SMOTE can be seen in Fig. 3.

Subsequently, all features underwent normalization, and One-Hot Encoding was employed to encode the labels of both datasets. Finally, a reshaping process was implemented for train-
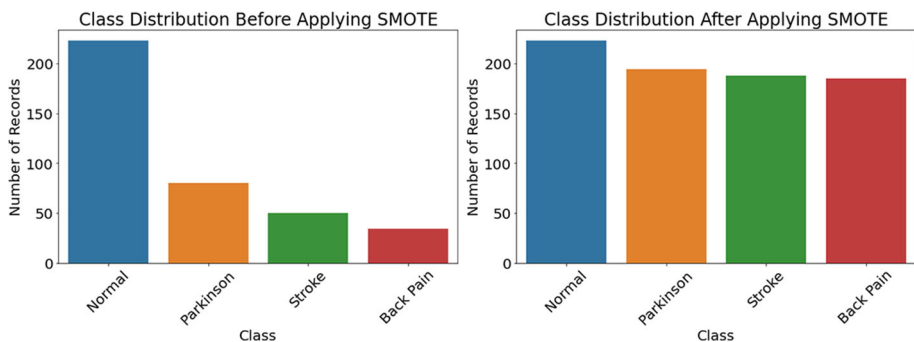

Fig. 3 Distribution of the five disease records in the KIMORE dataset before and after applying SMOTE

ing and testing data subsets to ensure their formats conform to the consistent data structure required by all utilized models. One-Hot Encoding is a common approach for multi-class datasets [68–71].

4.3 Hyper-parameters tuning

Deep learning poses a significant challenge in terms of model optimization. Despite the widespread success of deep networks across various domains, certain factors can hinder their overall effectiveness. These factors encompass the selection of appropriate model parameters, determining the architecture and feature representation, and identifying optimal weight and bias values [72]. In this research, we harnessed the capabilities of four deep learning algorithms, specifically LSTM, BiLSTM, CNN-LSTM, and CNN. Our central concern was the identification of the most suitable Hyper-parameters for these models. Our effort went beyond finding these four models best architecture and parameters. It encompassed the pursuit of Hyper-parameters capable of consistently delivering exceptional results across both datasets. The manual exploration of diverse configurations and subsequent evaluation to ascertain the optimal parameter settings is an onerous and time-consuming endeavor, especially when dealing with high-dimensional search spaces. Furthermore, it is crucial to recognize that finely tuned Hyper-parameter configurations, established by an experienced user, are intricately tied to the specific application. Consequently, a compelling need exists for automated Hyper-parameter optimization to alleviate computational demands and minimize user intervention. Utilizing the steepest gradient descent algorithm for optimizing Hyper-parameter configurations proves unsuitable [72]. Instead, prominent techniques such as Grid Search and Random Search have emerged to address this challenge effectively. Grid search is a brute-force approach to Hyper-parameter tuning. It evaluates all possible combinations of Hyper-parameter values [73], whereas Random Search adopts a more stochastic approach by exploring random subsets of Hyper-parameter combinations [74]. The choice between Grid Search and Random Search hinges significantly on the number of Hyper-parameters under consideration. Grid Search proves more convenient when the Hyper-parameter count is limited, whereas Random Search excels when dealing with a larger number of Hyper-parameters [74]. In our approach, we have adapted Random Search to configure the models architectures and determine the most favorable Hyper-parameters that consistently yield superior results across both datasets. Figure 2 [Model Tuning] Visualizes the architecture designing and Hyper-parameters-tuning phase.

The Random Search technique, as discussed in [74], was employed to select the model architecture and fine-tune Hyper-parameters across four distinct models. Due to computational resource limitations, 150 trials were conducted for all models. The primary objective of this endeavor was to optimize validation loss during the training process. The dataset was partitioned to achieve this, with 80% of the data allocated for training and validation purposes. Hyper-parameter tuning was performed on one dataset, and subsequently, the model was tested on another dataset. We explored four distinct combinations: (1) Tuning on the KIMORE dataset and training on the UI-PRMD dataset. (2) Tuning on the UI-PRMD dataset and training on the KIMORE dataset. (3) Initial tuning on the KIMORE dataset followed by a subsequent round of Hyper-parameter optimization on the UI-PRMD dataset. (4) Initial tuning on the UI-PRMD dataset followed by a subsequent round of Hyper-parameter optimization on the KIMORE dataset. Our findings led us to determine that the most favorable results were obtained through the second combination, where the model was tuned on the UI-PRMD dataset and subsequently tested on the KIMORE dataset. This section compre-

hensively explains the parameter values considered for investigation for each model. For the (LSTM) model, The Random Search Tuner was employed to ascertain the values for several critical variables to establish the most effective model architecture. These variables encompassed the following: the number of (LSTM) layers, number of (LSTM) units, dropout rate, learning rate (ranging from 0.0001 to 0.01), type of regularizer (l1, l2, or none), and its associated strength [75]. The regularization strength parameter, often denoted as lambda (λ), controls the amount of regularization applied to the model. A higher value of λ corresponds to more robust regularization, which penalizes larger weights more heavily, leading to a more sparse model. Conversely, a lower value of λ reduces the regularization effect, allowing the model to have larger weights.

A comprehensive set of Hyper-parameters was considered in the context of the (CNN-LSTM) model. These parameters encompassed several key aspects, including the count of convolutional filters, the kernel size, the quantity of (LSTM) units, the dropout rate, the learning rate, and the choice of regularization techniques. These techniques are applied to convolutional, dense, and (LSTM) layers, providing flexibility with options such as l1, l2, or none.

Similarly, for the (BiLSTM) model, an array of Hyper-parameters came into play. These parameters involved factors like the number of Bidirectional LSTM layers, the quantity of (LSTM) units, the dropout rate, the learning rate, and the selection of regularization techniques. Among the choices for regularization were options like l1, l2, or none.

Shifting the focus to the (CNN) model, the Hyper-parameters under consideration included the count of convolutional layers, the number of dense layers, convolutional units, dense units, the dropout rate, the application of regularization techniques to both convolutional and dense layers (with choices of l1, l2, or none), and the learning rate. These parameters were pivotal in shaping the model's architecture and behavior during training and evaluation.

4.4 Action classification

LSTM variants stand out in action classification owing to their robust memory retention capabilities, facilitating the effective capture and interpretation of temporal dependencies within sequential data-an essential aspect of action classification tasks. This advantage is well-established in natural language processing and time-series analysis, where LSTM architectures consistently outperform conventional models in processing sequential information. The study encompasses three investigated LSTM variants: LSTM, Bidirectional LSTM (BiLSTM), and Convolutional Neural Network - LSTM (CNN LSTM). All models process input data, comprising a 500-dimensional feature vector for the KIMORE dataset and a 330-dimensional feature vector for the UI-PRMD dataset. The output layer includes 20 classes for the UI-PRMD dataset and five classes for the KIMORE dataset, with these model selections being motivated by previous research studies that showcase the widespread adaptation of LSTM networks in various domains, such as human activity recognition [16–18], chronic pain detection [19], and exercise classification [76, 77]. Similarly, Bidirectional LSTM (BiLSTM) has found extensive application in studies related to human activity recognition [20–22], exercise classification [78, 79], and other relevant domains. Convolutional Neural Networks (CNN) have demonstrated remarkable results in human activity recognition [23–25] and exercise classification [80–84]. The hybrid CNN-LSTM model has also been utilized in both human activity recognition [26, 27, 51] and physical rehabilitation domains [34, 52, 53].

characteristics

4.4.1 Long short-term memory (LSTM)

Long Short-Term Memory (LSTM) [85] networks have emerged as a prominent and efficacious class of recurrent neural networks (RNNs) that can effectively model and process sequential data. In contrast to conventional RNNs, LSTMs possess the distinctive ability to mitigate the vanishing gradient problem through a complex gating mechanism, which facilitates the preservation of long-range dependencies. This is achieved by regulating the flow of information through a network of memory cells, each equipped with input, output, and forget gates. The (LSTM) architecture's adaptive gating mechanism enables it to capture and retain essential information over extended sequences, making it particularly well-suited for tasks involving temporal dynamics and sequential patterns. Moreover, the introduction of variants such as Bidirectional LSTM (BiLSTM) and Convolutional Neural Network - LSTM (CNN LSTM) has further extended the (LSTM) framework's capabilities, allowing for the exploitation of bidirectional context and spatial features, respectively.

The Hyper-parameter tuning process determined that the model parameters and architecture comprise a single (LSTM) layer with 320 units. A learning rate of 0.000501 was identified as the most effective value for the learning process. To mitigate the problem of over-fitting, we introduced a dropout rate of 0.263337 and implemented an L2 regularizer. Dropout layers were specifically employed to counteract over-fitting, which, although potentially increasing training time, significantly contributes to enhancing the model's generalization capability. Moreover, batch normalization can achieve model generalization, although dropout offers finer control over regularization strength [86]. Two dense layers further complemented this architecture, each comprising 940 units. Finally, the output layer utilized Softmax [87] activation due to its superiority over other activation functions.
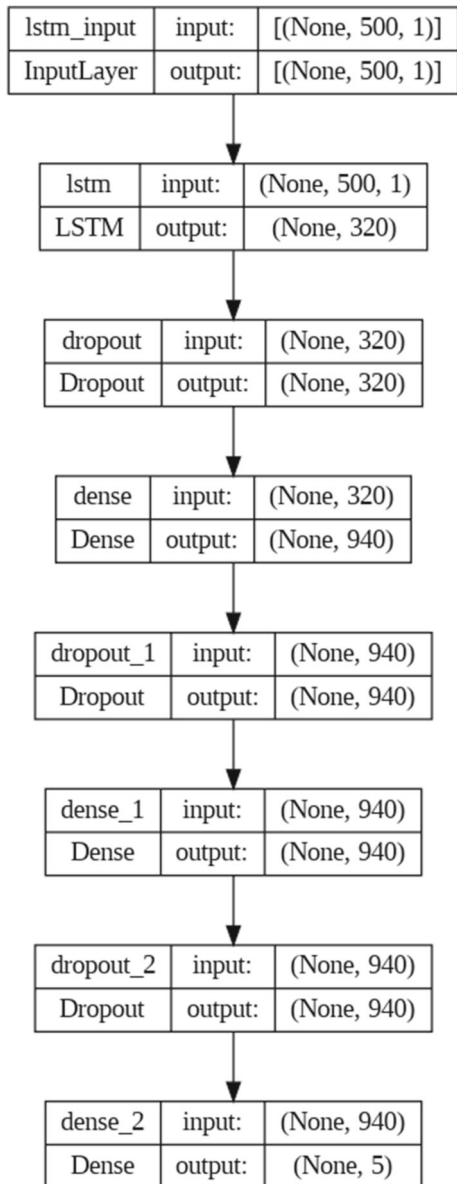
Figure 4 visually illustrates the proposed model's architecture, which was arrived at through a comprehensive exploration of various (LSTM) configurations and fine-tuning numerous model parameters via random search.

4.4.2 Bidirectional Long Short-Term Memory (BiLSTMs)

BiLSTMs, short for Bidirectional Long Short-Term Memory networks, introduce bidirectional processing by integrating two distinct (LSTM) layers that analyze input sequences in both forward and reverse directions [88]. This unique design allows the network to capture past and future context concurrently, significantly enhancing its ability to discern intricate temporal relationships within sequential data. Integrating information from both directions makes BiLSTMs exceptionally proficient in tasks demanding a profound understanding of context, such as natural language comprehension and sentiment analysis. This bidirectional processing effectively mitigates the limitations associated with unidirectional RNNs, which often struggle with capturing long-range dependencies and mitigating the vanishing gradient problem. Consequently, BiLSTMs exhibit superior performance on tasks requiring nuanced sequential information processing.

Following Hyper-parameter tuning, the parameters and architecture were determined to consist of two (BiLSTM) layers, each comprising 271 units. Additionally, a learning rate of 0.001014 was identified as the value for the learning process. A dropout rate of 0.3 was used to mitigate the over-fitting problem further. The model also incorporated four dense layers, each containing 927 units with ReLU activation, except for the output layer, which utilized softmax activation. The architectural layout presented in Fig. 5 was devised after evaluating various (BiLSTM) implementations and fine-tuning multiple parameters using a random search approach.
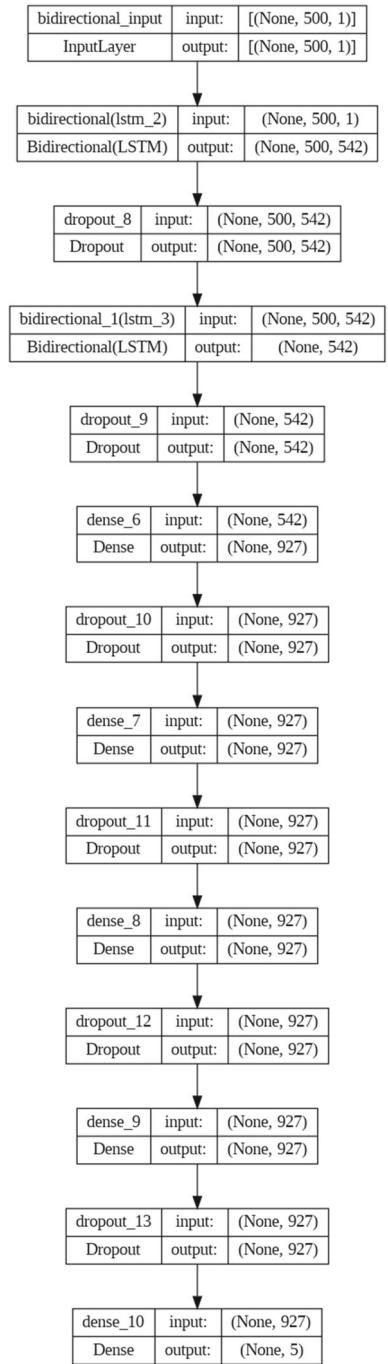
Fig. 4 The architecture of
(LSTM) Model after applying
Random Search for Hyper-tuning

lstm_input	input:	[(None, 500, 1)]
InputLayer	output:	[(None, 500, 1)]

lstm	input:	(None, 500, 1)
LSTM	output:	(None, 320)

dropout	input:	(None, 320)
Dropout	output:	(None, 320)

dense	input:	(None, 320)
Dense	output:	(None, 940)

dropout_1	input:	(None, 940)
Dropout	output:	(None, 940)

dense_1	input:	(None, 940)
Dense	output:	(None, 940)

dropout_2	input:	(None, 940)
Dropout	output:	(None, 940)

dense_2	input:	(None, 940)
Dense	output:	(None, 5)

4.4.3 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN), initially pioneered by Yann LeCun et al., [89], represent a groundbreaking innovation in deep learning and computer vision. Designed initially for image recognition tasks, (CNNs) are inspired by the hierarchical feature extraction mechanisms of the human visual system, replicating the autonomous process through which the human brain distills relevant features from input data. Comprising convolutional layers for feature extraction followed by pooling layers for spatial dimension reduction, CNNs exhibit

Fig. 5 The architecture of (BiLSTM) Model after applying Random Search for Hyper-tuning

bidirectional_input	input:	[(None, 500, 1)]
InputLayer	output:	[(None, 500, 1)]

bidirectional(lstm_2)	input:	(None, 500, 1)
Bidirectional(LSTM)	output:	(None, 500, 542)

dropout_8	input:	(None, 500, 542)
Dropout	output:	(None, 500, 542)

bidirectional_1(lstm_3)	input:	(None, 500, 542)
Bidirectional(LSTM)	output:	(None, 542)

dropout_9	input:	(None, 542)
Dropout	output:	(None, 542)

dense_6	input:	(None, 542)
Dense	output:	(None, 927)

dropout_10	input:	(None, 927)
Dropout	output:	(None, 927)

dense_7	input:	(None, 927)
Dense	output:	(None, 927)

dropout_11	input:	(None, 927)
Dropout	output:	(None, 927)

dense_8	input:	(None, 927)
Dense	output:	(None, 927)

dropout_12	input:	(None, 927)
Dropout	output:	(None, 927)

dense_9	input:	(None, 927)
Dense	output:	(None, 927)

dropout_13	input:	(None, 927)
Dropout	output:	(None, 927)

dense_10	input:	(None, 927)
Dense	output:	(None, 5)

a hierarchical feature learning capability that enables them to discern intricate patterns and features within datasets.

In the context of exercise classification, the superior performance of the CNN model can be attributed to its adeptness in extracting hierarchical spatial features from skeletal time-series data, such as that captured by Kinect cameras during exercise execution. The model's robustness to spatial transformations and its capacity for hierarchical feature abstraction contribute to its remarkable accuracy rates in discerning various exercises. Integrating these architectural features within the CNN framework enhances its efficacy for exercise classification tasks, showcasing its prowess in leveraging spatial relationships for accurate and robust recognition of distinct movement patterns.

Moreover, CNNs have demonstrated remarkable adaptability and versatility beyond image analysis, expanding their application to diverse domains. Notably, CNNs are extensively utilized in disease classification within physical rehabilitation [90–93]. This highlights the broad impact of CNNs in leveraging sensor data for tasks ranging from human activity recognition to disease classification.
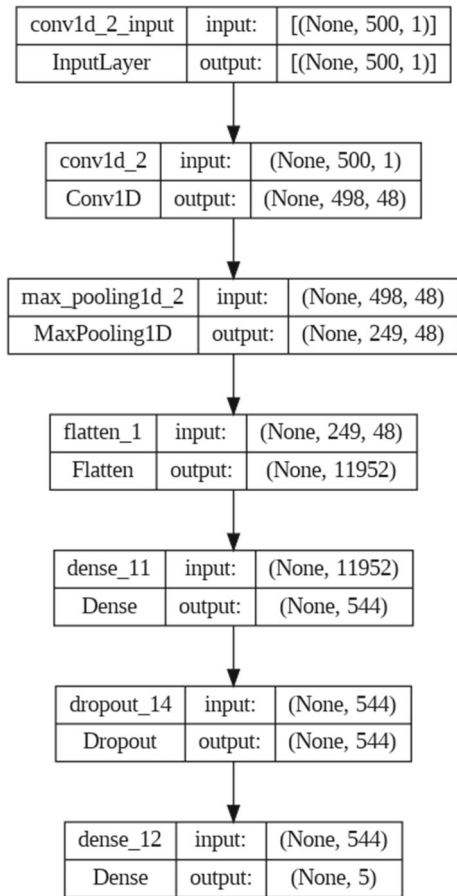
Through Hyper-parameter tuning, the parameters and architecture were determined to comprise two convolutional layers, each housing 48 convolutional units with a 3x3 filter size. A Max-Pooling Layer and a Flatten Layer followed this. Additionally, a dense layer consisting of 544 units with ReLU activation was incorporated, followed by a dropout layer with a rate of 0.2 to address potential over-fitting. A learning rate of 0.0025284 was identified as the value for the learning process, and the model concluded with an output layer featuring softmax activation. Figure 6 visually represents the architecture of the proposed model. This architectural design was adopted after experimenting with various (CNN) implementations and fine-tuning multiple model parameters using a random search approach.

4.4.4 CNN-LSTM

The (CNN-LSTM) architecture seamlessly integrates Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, harnessing both spatial and temporal information synchronously [94]. This fusion of CNNs for spatial feature extraction and LSTMs for temporal pattern capture excels in video analysis and action recognition tasks. Its significance is particularly pronounced in addressing challenges involving visual content and its temporal evolution. This architecture finds versatile applications in fields like video surveillance, medical imaging, and natural language processing, enriching insights from sequential data by considering spatiotemporal relationships. In essence, (CNN-LSTM) offers an efficient amalgamation of CNNs and LSTMs, providing a robust approach for deciphering intricate data.

This study introduces a hybrid (CNN-LSTM) model for classification purposes. Here, (CNN) serves as the feature extractor, while (LSTM) takes charge of classification, as illustrated in Fig. 7. The model parameters and architecture commence with a convolutional layer comprising 128 filters, featuring a kernel size of 8, along with L2 and bias regularizers set at 0.002276. A max-pooling layer and a dropout layer with a rate of 0.2 follow this. Subsequently, an (LSTM) layer with 256 units is employed, succeeded by a flattened layer. Two dense layers are added, each with 525 units and ReLU activation, complemented by a dropout rate of 0.2. The learning rate identified for the learning process is 0.000777. Finally, the output layer utilizes softmax activation. The architectural layout of the model is visually depicted in Fig. 7. This architecture was adopted after conducting experiments with various (CNN-LSTM) implementations and fine-tuning multiple model parameters using a random search approach.
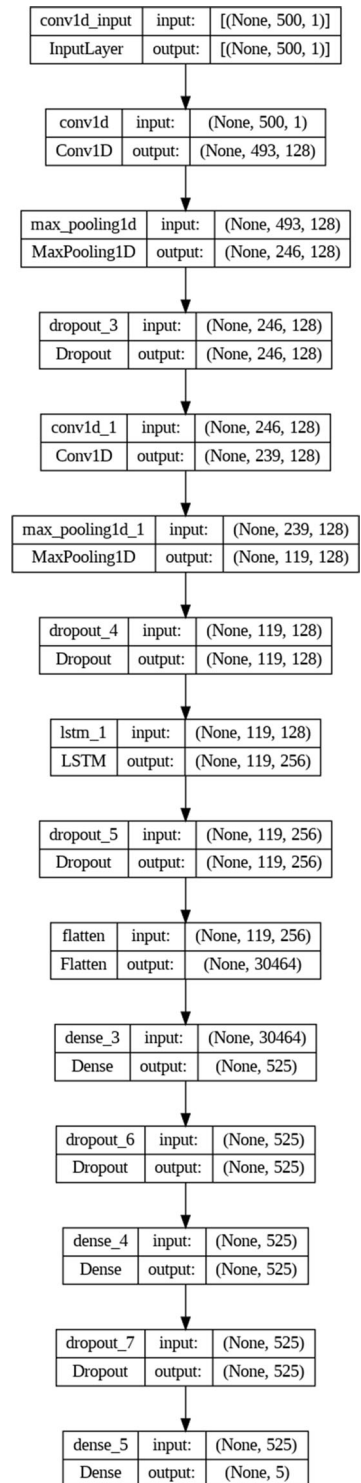
Fig. 6 The architecture of (CNN) Model after applying a Random Search for Hyper-tuning

conv1d_2_input	input:	[(None, 500, 1)]
InputLayer	output:	[(None, 500, 1)]

conv1d_2	input:	(None, 500, 1)
Conv1D	output:	(None, 498, 48)

max_pooling1d_2	input:	(None, 498, 48)
MaxPooling1D	output:	(None, 249, 48)

flatten_1	input:	(None, 249, 48)
Flatten	output:	(None, 11952)

dense_11	input:	(None, 11952)
Dense	output:	(None, 544)

dropout_14	input:	(None, 544)
Dropout	output:	(None, 544)

dense_12	input:	(None, 544)
Dense	output:	(None, 5)

4.5 Model deployment and monitoring

The deployment of a model in a production environment constitutes a pivotal phase in its operational integration. Consequently, adopting model deployment with MLOps, as outlined in [95], becomes necessary. To facilitate model consumption, the design of an API using Python frameworks, notably Flask, has been considered, drawing insights from various articles [96, 97], particularly those focusing on disease classifications involving Flask development and its integration with machine learning models. The selection of Python is motivated by its pre-defined packages, such as Tensorflow and Pandas [95]. The model can be deployed in the production environment for its usage. Various cloud platforms, such as Amazon Web Service (AWS), Microsoft Azure, or Google Cloud, can be utilized to deploy the model. In this MLOps-driven deployment, an automated pipeline takes charge of continuous monitoring, incorporating automated testing, validation services, and responsive actions to maintain model accuracy. The MLOps architecture facilitates a seamless machine learning life-cycle, ensuring adaptive retraining of the model and consistent alignment with predefined performance standards. Automated retraining can be utilized according to a predefined schedule, focusing on ensuring thorough documentation and logging of the entire process.

Fig. 7 The architecture of (CNN-LSTM) Model after applying a Random Search for Hyper-tuning

conv1d_input	input:	[(None, 500, 1)]
InputLayer	output:	[(None, 500, 1)]

conv1d	input:	(None, 500, 1)
Conv1D	output:	(None, 493, 128)

max_pooling1d	input:	(None, 493, 128)
MaxPooling1D	output:	(None, 246, 128)

dropout_3	input:	(None, 246, 128)
Dropout	output:	(None, 246, 128)

conv1d_1	input:	(None, 246, 128)
Conv1D	output:	(None, 239, 128)

max_pooling1d_1	input:	(None, 239, 128)
MaxPooling1D	output:	(None, 119, 128)

dropout_4	input:	(None, 119, 128)
Dropout	output:	(None, 119, 128)

lstm_1	input:	(None, 119, 128)
LSTM	output:	(None, 119, 256)

dropout_5	input:	(None, 119, 256)
Dropout	output:	(None, 119, 256)

flatten	input:	(None, 119, 256)
Flatten	output:	(None, 30464)

dense_3	input:	(None, 30464)
Dense	output:	(None, 525)

dropout_6	input:	(None, 525)
Dropout	output:	(None, 525)

dense_4	input:	(None, 525)
Dense	output:	(None, 525)

dropout_7	input:	(None, 525)
Dropout	output:	(None, 525)

dense_5	input:	(None, 525)
Dense	output:	(None, 5)

This includes logging predictions, inputs, and outputs for each inference and facilitating auditing and debugging processes.

5 Experiments

Two experiments were conducted in this research. The First Experiment was conducted to find the best exercise classification algorithm across both datasets. The Second experiment was conducted to classify different diseases from patients while performing the same five exercises in the KIMORE dataset.

5.1 Experiment 1: finding the best algorithm for exercise classification

The experiment assessed and contrasted the performance of distinct (LSTM) variations as long as (CNN) models. The dataset was partitioned into training and testing subsets to an 80-20 ratio for training and testing. Furthermore, the training data was divided within the training subset into an 80-20 ratio for training and validation data, which was utilized for model training. This splitting employed the same number of folds as [35]. A 5-fold cross-validation approach was adopted to mitigate over-fitting, maintaining a consistent batch size of 32 for all models. This batch size choice aligns with the methodology employed in [29]. All models employed an identical loss function of categorical cross-entropy and were trained for 450 epochs. An early stopping technique was also applied. All experiments were run on the same machine with 15GB of GPU. Early stopping is a regularization technique used in machine learning to prevent over-fitting. It stops a model's training process when it starts over-fitting the training data [98]. The Early Stopping technique was implemented to monitor validation loss during training with a patient of 30 epochs. All models were trained simultaneously on the same machine using the same computational resources. Adam [99] (short for Adaptive Moment Estimation) was also used as an optimization algorithm for training machine learning models, particularly neural networks. It is an extension of stochastic gradient descent (SGD) and is designed to optimize the learning process by adapting the learning rates of individual model parameters. Figure 2 [Data Splitting] visualizes the data splitting phase. While [Training] Visualizes the training process and Hyper-parameters optimization.

5.2 Experiment 2: finding the best algorithm for disease classification

This investigation assessed and juxtaposed the efficacy of various (LSTM) variations and (CNN) models in disease classification utilizing the KIMORE dataset. Following Experiment 1, we maintained adherence to the established experimental protocol. This involved employing a consistent 5-fold cross-validation approach, implementing an identical early stopping callback throughout training for 450 epochs, and utilizing the same categorical cross-entropy loss function with a batch size 32.

To rectify class imbalance, we implemented the SMOTE technique for oversampling.

5.3 Evaluation metrics

The selection of specific evaluation metrics is a crucial aspect of assessing the performance of machine learning models, and it involves a thoughtful consideration of the dataset's char-

acteristics and the study's goals. In this work, we opted for a set of well-established metrics, namely loss, accuracy, precision, recall, and F1-score, to comprehensively evaluate the proposed models.

Our choice is motivated by the balanced nature of the datasets. While accuracy is a commonly used metric that provides a general overview of overall model correctness, its effectiveness can be influenced by the class distribution. In a balanced dataset, accuracy is a suitable metric, and its inclusion ensures a broad assessment of model performance.

On the other hand, precision, recall, and F1-score are particularly relevant in scenarios where a balanced consideration of false positives and false negatives is imperative. Precision assesses the accuracy of positive predictions, recall measures the ability to capture all relevant instances, and F1 score balances precision and recall. These metrics become especially significant when the consequences of false positives and false negatives differ in importance.

Furthermore, incorporating loss as an evaluation metric offers insights into the convergence and optimization process during training. Minimizing the loss is fundamental to achieving accurate predictions, making it a valuable metric for assessing the training dynamics.

In alignment with related work utilizing accuracy and F1-score, our choice of metrics aims to provide a robust evaluation framework that considers various aspects of model performance. This approach ensures a nuanced understanding of the proposed models effectiveness in capturing global correctness and the balance between precision and recall, ultimately contributing to our evaluation process's scientific rigor and completeness.

6 Results

The Results section is organized into four main parts: Exercise Classification (Section 6.1), a Comparison of State-of-the-art Methods (Section 6.2), Disease Classification (Section 6.3), and an overview of Real-world Applications in (Section 6.4).

6.1 Exercise classification

The (CNN) model demonstrates the most efficient convergence in training, requiring the fewest iterations to achieve lower loss values during the training process on both datasets. The exercise classification results are further divided into two subsections, one for each dataset.

6.1.1 KIMORE

In our experimental investigation of the KIMORE dataset, the Hyper-tuned (BiLSTM) model emerged as the top-performing candidate, outperforming other models across all five critical evaluation metrics. It achieved remarkable results with a mean accuracy of 93.08%, precision of 93.07%, recall of 93.96%, an F1-score of 91.79%, and a loss value of 0.2860. Following closely, the (CNN-LSTM) and Bi-LSTM models also delivered commendable results, boasting accuracy rates of 91.79% and 87.18%, respectively. Notably, the Bi-LSTM model exhibited a shorter mean training duration compared to the (CNN-LSTM) model, with the (CNN-LSTM) consuming the highest duration of 7.88 minutes on average per fold. In contrast, the (CNN) model incurred the shortest training time, averaging 2.3 minutes. It's worth mentioning that the (CNN) model, while computationally expensive and resource-intensive

due to its large number of trainable parameters (6,505,349), benefits from early stopping, requiring the fewest iterations to converge, as demonstrated in Fig. 8.

However, the (LSTM) model displayed a comparatively less impressive performance in this experimental context, yielding an average accuracy rate of 82.32%. This discrepancy may be attributed to the intrinsic nature of (LSTM) models, which thrive when provided with a substantial number of time steps. In our experiment, we deliberately constrained the time steps to a single record.

Simultaneously, Fig. 8 presents a graphical representation of the precision metrics in [b], the F1-score in [c], and the recall metric in [d]. These visual presentations offer valuable insights into the progressive convergence patterns exhibited by the models during the training phase when utilizing the validation dataset.
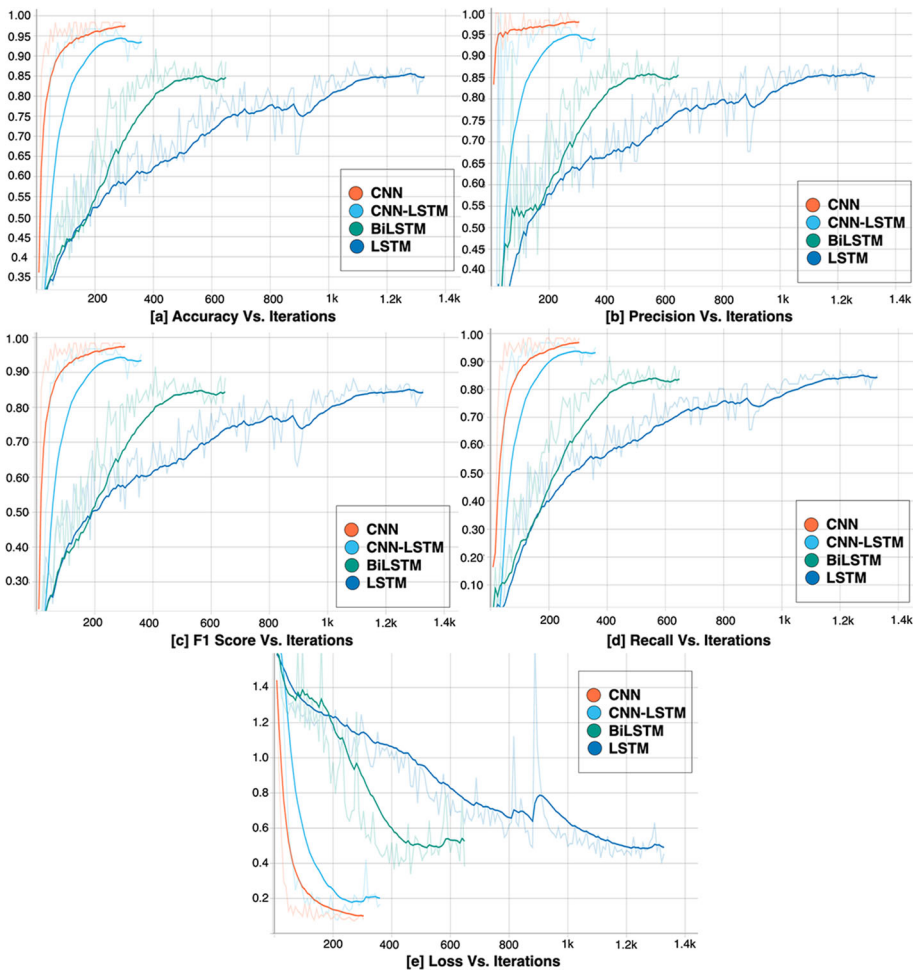
Fig. 8 Iterations results on the KIMORE dataset, featuring five subgraphs illustrating the performance comparison of four models across different evaluation metrics: [a] Accuracy, [b] Precision, [c] F1-score, [d] Recall, and [e] Loss

Table 4 The Mean Performance Metrics of Various Models on the KIMORE dataset, Alongside the Duration Time for Each Fold (Measured in Minutes)

Algorithm	Accuracy	Precision	Recall	F1-Score	Loss	Training Time
LSTM	82.31%	82.40%	84.65%	79.49%	0.5519	2.45
BiLSTM	87.18%	87.34%	88.75%	86.67%	0.4608	5.62
CNN-LSTM	91.79%	91.79%	92.65%	90.77%	0.4139	7.88
CNN	**93.08%**	**93.07%**	**93.96%**	**91.79%**	**0.2860**	**2.30**

Additionally, as depicted in Fig. 8 [e], we present graphical representations of the loss values for all four models during the final fold, which is the lowest-scoring fold among the five, on the KIMORE dataset. The corresponding number of training iterations is also included in the visualization. Furthermore, Fig. 8 [a] provides a visual overview of the recorded accuracy metrics.

Table 4 presents a comprehensive comparison of mean results obtained by each of the four models, employing diverse evaluation metrics during the rigorous testing phase. Correspondingly, Fig. 9 supplements this analysis by providing a whisker box chart, which highlights the distribution of performance metrics, including minimum, maximum, mean, and interquartile range (IQR) values for the four models across the four evaluation metrics. This visual representation offers a detailed overview of the performance variations and statistical distribution of results for each model across all 5 folds.

6.1.2 UI-PRMD

For the second dataset, UI-PRMD, the (CNN) model showcased exceptional performance across a comprehensive spectrum of evaluation metrics, encompassing Accuracy, Precision, Recall, F1-Score, and Loss. The (CNN) model also exhibited remarkable efficiency, with an average training duration of just 2.15 minutes across five-folds, thanks to its reduced iteration requirements. This model achieved outstanding accuracy (99.70%), precision (99.70%), recall (99.95%), and F1-score (99.70%), along with a minimized loss value of 0.0122.
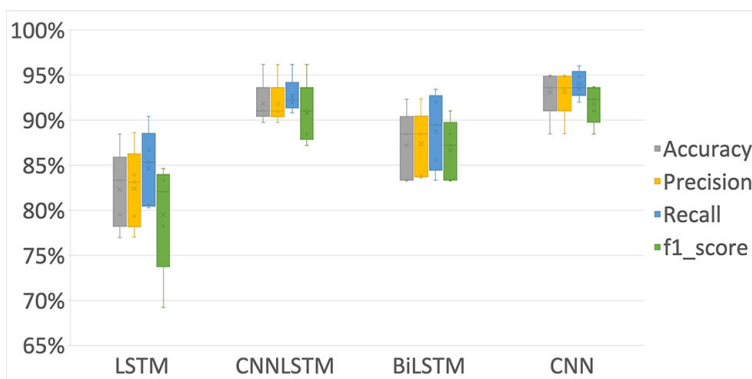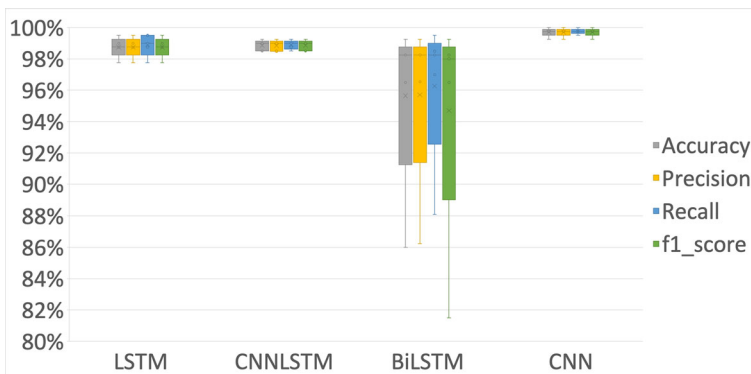

Fig. 9 Key performance metrics, encompassing accuracy, precision, recall, and F1 score are presented through whisker boxes for the 4 models. This visual representation underscores optimal outcomes on the KIMORE dataset for exercise classification following a thorough 5-fold cross-validation

Table 5 The Results of the different models on the UI-PRMD dataset

Algorithm	Accuracy	Precision	Recall	F1-Score	Loss	Training Time
BiLSTM	95.65%	95.71%	96.27%	94.70%	0.1434	13.45
LSTM	98.75%	98.75%	98.90%	98.75%	0.0626	4.18
CNN-LSTM	98.85%	98.85%	98.90%	98.85%	0.0676	8.54
CNN	**99.70%**	**99.70%**	**99.75%**	**99.70%**	**0.0122**	**2.15**

In the second position, the (CNN-LSTM) model achieved commendable results, with an accuracy of 98.85%, precision of 98.85%, recall of 98.90%, F1-score of 98.90%, and a loss value of 0.0676. However, it necessitated a relatively extended training duration, with an average training time of 8.45 minutes. The (CNN-LSTM) model, despite its complexity with 16,799,497 parameters, demonstrated competitive performance and required slightly fewer iterations compared to other algorithms, as shown in Fig. 8.

Conversely, both the Bi-LSTM and (LSTM) models delivered acceptable results, with accuracy rates of 96.65% and 98.75%, respectively. Notably, the (LSTM) model exhibited significantly shorter training times than Bi-LSTM, with an average of 4.18 minutes versus 13.45 minutes, making it nearly three times faster. These outcomes emphasize the superiority of the (CNN) model, the competitive performance of the (CNN-LSTM) model, and the limitations observed in the (LSTM) model within this specific experimental context.

Furthermore, Table 5 compares the mean results attained by each of the three models, utilizing various evaluation metrics during the testing phase. In addition, Fig. 10 presents an additional whisker box chart, illustrating the results of the models on the UI-PRMD dataset. As depicted in the figure, (CNN) surpassed all other models in performance and exhibited the least variance across all five folds.

6.2 Comparison of state-of-the-art

This section provides a comparison of the current state-of-the-art methodologies applied to the KIMORE and UI-PRMD datasets for the classification of exercises.
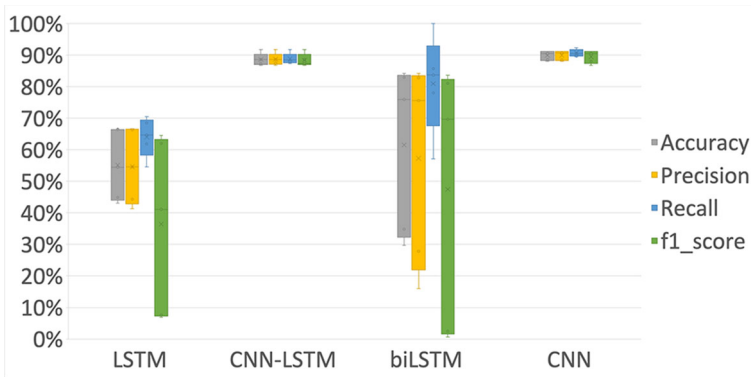

Fig. 10 Key performance metrics, encompassing accuracy, precision, recall, and F1 score are presented through whisker boxes for the four models. This visual representation underscores optimal outcomes on the UI-PRMD dataset following a thorough 5-fold cross-validation

Table 6 Comparison of our Tuned-CNN model with state-of-the-art methods on the KIMORE dataset

Algorithm	Accuracy
Ensemble-based Graph Convolutional Network (EGCN) [39]	80.1%
3D Convolution Neural Network (3D-CNN) [41]	90.57%
Many-to-Many model with density map output [35]	92.33%
Our Tuned-CNN	**93.08%**

The Tuned-CNN model demonstrates remarkable performance when contrasted with previous studies conducted on the same dataset (KIMORE), achieving an impressive accuracy rate of 93.08%. As per the extent of our investigation, this signifies a substantial enhancement over the state-of-the-art methods reported by Abedi et al., who employed LSTM, and a 1D-CNN followed by a fully connected neural network, showing improvements of 0.75%. Barzegar et al.,'s utilization of a 3D-CNN is also surpassed by 2.51% [35]. A comparative analysis of results obtained from prior research endeavors and the mean value of our highest-scoring model is presented in Table 6.

In Table 7, a comprehensive comparison of our proposed models with other relevant works on the UI-PRMD dataset is presented. Notably, our Tuned-CNN model achieves an exceptional accuracy of 99.7%, outperforming existing state-of-art methods such as Ensemble-based Graph Convolutional Network (EGCN) [39], Graph Convolutional Siamese Network [37] by 0.5%, and Extra Tree Classifier after incorporating the FCBF feature ranking technique [100] by 0.1%. This comparison underscores the superior performance of our tuned CNN model in accurately classifying exercises on the UI-PRMD dataset.

6.3 Disease Classification

Our primary objective shifted from exercise classification to disease identification during patient exercise in this experiment. Rather than conducting Hyper-parameter tuning to create entirely new models, we retained the original models architecture and Hyper-parameters. However, the results yielded less promising outcomes, primarily because these models were initially designed and optimized for exercise classification, not disease identification. Although we used the same dataset, the fundamental goal of this experiment significantly differed from the previous one. Consequently, while we expected the tuned models to demonstrate excellent performance on this specific dataset, their performance on other datasets remained uncertain.

Table 7 Comparison of our Tuned-CNN model with state-of-the-art methods on the UI-PRMD dataset

Algorithm	Accuracy
Ensemble-based Graph Convolutional Network (EGCN) [39]	86.9%
Graph Convolutional Siamese Network [37]	99.2%
FCBF - Extra Tree [100]	99.6%
Our Tuned-CNN	**99.7%**

Fig. 11 Key performance metrics, encompassing accuracy, precision, recall, and F1 score are presented through whisker boxes for the four models. This visual representation underscores optimal outcomes on the KIMORE dataset for disease classification following a thorough 5-fold cross-validation

Therefore, we decided not to perform model tuning on the new dataset to avoid over-fitting. Furthermore, it's important to note that using the same architecture on other available datasets for disease classification is not feasible as such datasets are scarce.

We present the results as a whisker box chart to provide a comprehensive overview of the model's performance consistency across different folds in disease classification. This chart, visualized in Fig. 11, displays the distribution of accuracy values across the five folds. Each whisker box represents the interquartile range (IQR) of accuracy, with the central line denoting the median accuracy. Outliers, if any, are also highlighted, offering insights into the variability of model performance. Examining this graphical representation enhances our understanding of the stability and reliability of the proposed models over multiple folds, contributing valuable information for assessing the robustness of the disease classification system. Detailed statistical measures such as mean, standard deviation, and confidence intervals further complement the visual interpretation, comprehensively evaluating the models' accuracy across different folds.

Once again, the (CNN) model exhibited superior performance, outperforming all other algorithms, followed closely by CNN-LSTM, with only a tiny margin of 1.26%. Specifically, the (CNN) model achieved an accuracy of 89.87%, while (CNN-LSTM) scored 88.61%. Notably, the (CNN) model also boasted significantly faster training times, approximately 2.8 times faster than CNN-LSTM. In contrast, the Bi-LSTM model scored an average accuracy of 61.52%, while the (LSTM) model struggled with just 55.06%. Table 8 lists the median results across all metrics during the five folds.

Table 8 The Average Performance Metrics of Various Models on the KIMORE dataset for **Disease Classification**

Algorithm	Accuracy	Precision	Recall	F1-Score	Loss	Training Time
LSTM	55.06%	54.63%	64.02%	36.64%	1.2521	**3.98**
BiLSTM	61.52%	57.30%	80.91%	47.47%	0.9633	9.07
CNN-LSTM	88.61%	88.64%	88.80%	88.35%	0.8937	20.89
CNN	**89.87**%	**89.91'**%	**90.63**%	**89.49**%	**0.48286**	7.32

It is essential to emphasize that while the results may not be outstanding, this represents an initial experiment. No model or Hyper-parameter tuning was performed due to the constraints imposed by the limited availability of datasets encompassing multiple diseases. Within the scope of our research, other datasets for disease classification typically comprise only binary classes, namely Normal and Affected.

In the context of disease classification, the performance evaluation of our proposed models includes a detailed analysis based on the confusion matrix and ANOVA analysis. The confusion matrix results after the 5th fold, which is considered the best fold, provide a comprehensive visualization of the model's classification performance across four distinct classes: normal, Parkinson's disease, back pain, and stroke. This crucial information is graphically represented in Fig. 12, which illustrates the model's classification outcomes for each class. The visualization aids in interpreting the effectiveness of the models in correctly categorizing patients into the respective disease classes based on the performed exercises. This visual representation can give further insights into precision, recall, and F1-score metrics, contributing to a holistic understanding of the models disease classification capabilities.

ANOVA (Analysis of Variance) was conducted to assess the statistical differences in the disease classification performance among four models-Bi-LSTM, CNN, CNN-LSTM, and LSTM-utilized on the KIMORE dataset. Table 9 presents the ANOVA results, showcasing
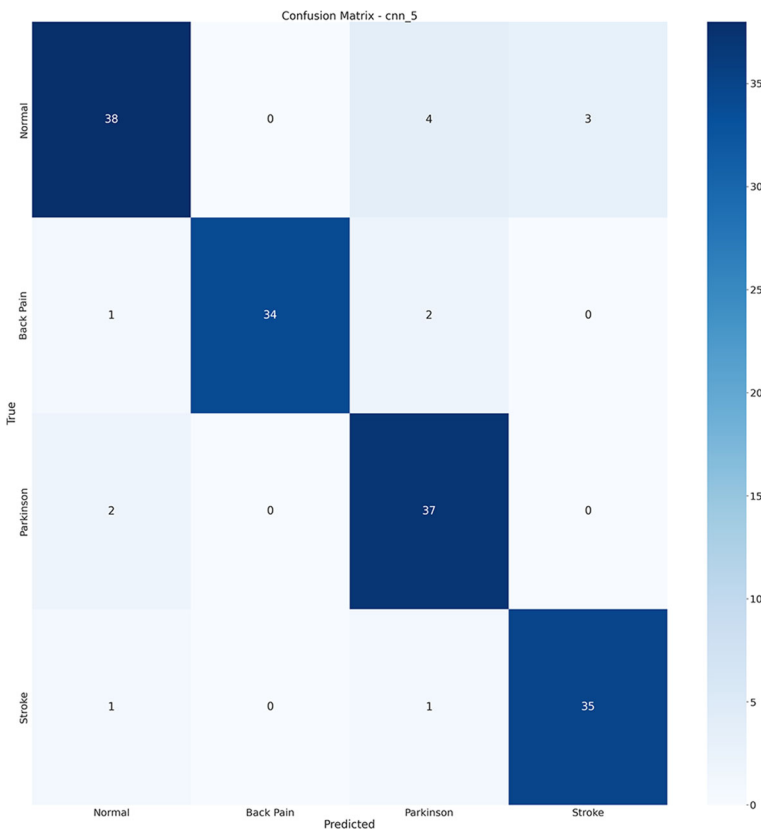
Fig. 12 Confusion Matrix illustrating disease classification results obtained from the (CNN) model after the 5th fold

Table 9 ANOVA results comparing four models architectures on the KIMORE dataset for disease classifcation

Models	Count	Sum	Average	Variance		
Bi-LSTM	5	3.08	61.52%	0.07255247		
CNN	5	4.49	89.87%	0.000220317		
CNN-LSTM	5	4.43	88.61%	0.000380548		
LSTM	5	2.75	55.06%	0.012678259		
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.489751272	3	0.163250424	7.607940975	0.002210221	3.238871517
Within Groups	0.343326373	16	0.021457898			
Total	0.833077645	19				

key parameters such as count, sum, average, and variance for each model. The top section details the model-wise statistical metrics. In contrast, the lower section delineates the source of variation, Sums of Squares (SS), Degrees of Freedom (df), Mean Squares (MS), F-ratio (F), P-value, and the Critical F value (F crit). The F-ratio is particularly noteworthy as it indicates the ratio of between-group variance to within-group variance. In this context, a statistically significant F-ratio (7.61) with a corresponding low P-value (0.0022) suggests significant differences in disease classification performance among the models.

6.4 Real-world applications of the proposed model

The adoption of advanced machine learning technologies in physical rehabilitation not only marks a significant advance in healthcare but also contributes to sustainability in the medical sector. By leveraging models like 1D CNN, LSTM, BiLSTM, and CNN-LSTM, rehabilitation processes become more efficient and personalized, reducing the need for frequent physical visits and thus, the carbon footprint associated with healthcare delivery. This approach supports the Sustainable Development Goal (SDG) 3, aimed at ensuring health and well-being for all, by making rehabilitation more accessible and reducing inequalities in health services. Moreover, the emphasis on Human Activity Recognition (HAR) and the use of digital and virtual reality tools in rehabilitation align with the broader goals of sustainable healthcare by optimizing resource use and enhancing the quality of care. These innovations not only improve patient outcomes but also align with global efforts towards universal health coverage and sustainability in healthcare, showcasing a commitment to a future where healthcare is both advanced and environmentally responsible.

6.4.1 Case study 1: remote patient monitoring

Enabling healthcare providers to monitor patients' rehabilitation progress remotely, especially crucial for patients who cannot frequently visit healthcare facilities due to distance, cost, or mobility issues.

6.4.2 Case study 2: physical therapy and rehabilitation

Automating the assessment of patients' performance during physical therapy sessions, providing real-time feedback to both patients and therapists. This can enhance the recovery.

6.4.3 Case study 3: home-based rehabilitation

Offering a solution for patients to perform guided exercises at home with virtual supervision, increasing the accessibility and frequency of rehabilitation sessions. With VR, patients can engage in gamified rehabilitation exercises, making home-based therapy more enjoyable and effective. AR can overlay instructional content on the patient's environment, guiding them through exercises correctly.

6.4.4 Case study 4: performance assessment in sports

Assisting coaches and athletes in analyzing performance for training effectiveness, injury prevention, and technique improvement by providing detailed insights into each movement's execution.

6.4.5 Case study 5: elderly care and fall prevention

Monitoring the elderly for sudden movements or falls and assessing their physical activity levels to customize fall prevention exercises and improve overall mobility and independence.

6.4.6 Case study 6: interactive fitness applications

Integrating with fitness apps and devices to provide users with feedback on their workout routines, ensuring exercises are done correctly to maximize benefits and reduce the risk of injury. Fitness apps can leverage AR to project digital trainers into the user's space, providing personalized workout guidance, while VR can immerse users in engaging, virtual landscapes to enhance the workout experience.

6.4.7 Case study 7: educational tools for physiotherapy

The proposed model can be used to disease classification and diagnosis, especially for conditions like Parkinson's, by analyzing movement data for early detection and personalized care plans.

6.4.8 Case study 8: rehabilitation research

This model can be particularly effective in identifying patterns and anomalies in physical movements associated with disorders such as Parkinson's disease and various gait abnormalities. Through detailed analysis of exercise performance data, the model aids in the early detection of these conditions, offering a significant improvement in diagnostic accuracy and the potential for early intervention. This application underscores the model's utility in not just rehabilitation but also in the proactive management and diagnosis of movement-related health issues, showcasing a versatile approach to healthcare.

6.4.9 Case study 9: disease classification and diagnosis through movement analysis

The proposed model can be used in disease classification and diagnosis, especially for conditions like Parkinson's, by analyzing movement data for early detection and personalized care plans.

7 Discussion

Using the Random Search technique to design diverse network architectures and fine-tune Hyper-parameters significantly impacted model performance. Our proposed Tuned-CNN model achieved state-of-the-art results on the KIMORE and UI-PRMD datasets, showcasing exceptional performance across different domains. Similarly, the Tuned-CNN-LSTM model demonstrated impressive and competitive performance on both datasets, highlighting its adaptability and effectiveness.

Interestingly, the (CNN) despite its complexity and a larger number of trainable parameters, exhibited efficient learning and the shortest training times. This efficiency resulted from requiring the fewest iterations and benefiting from early stopping. On the other hand, the Tuned-CNN-LSTM and Tuned-BiLSTM models demanded greater computational resources,

longer training times, and higher resource consumption. Notably, the (CNN) model achieved a remarkable accuracy of 99.70% on the UI-PRMD dataset and 93.08% accuracy on the KIMORE dataset for exercise classification. Finally, we recommend prioritizing using our Tuned-BiLSTM architecture over the Tuned-LSTM model.

The (CNN) and (CNN-LSTM) models also demonstrated impressive performance in disease classification on the KIMORE dataset, achieving accuracy rates of 89.87% and 88.61%, respectively.

Conversely, the (LSTM) model demonstrated consistent but sub-optimal results across both datasets. This performance discrepancy may be attributed to the algorithms inherent characteristics, particularly the LSTM's need for more timesteps. During our feature engineering phase, we generated only one timestep per record, potentially limiting the LSTM's capacity to excel.

The results highlighted the exceptional performance of the tuned (CNN) and (CNN-LSTM) models, consistently outperforming others across all evaluation metrics. The (CNN) model achieved remarkable accuracy scores of 93.08% and 99.7% on the KIMORE and UI-PRMD datasets, respectively. Notably, this performance surpassed the previous research on the same dataset by 0.75% on the KIMORE dataset and 0.1% on the UI-PRMD dataset. Our tuned (CNN) and (CNN-LSTM) models yielded very close results for disease classification on the KIMORE dataset, achieving an accuracy rate of nearly 89.9% and 88.61%, respectively. Specifically, the (CNN) model outperformed the (CNN-LSTM) model by approximately 1.26%, demonstrating significantly faster training times.

Despite the promising outcomes of our study, it is essential to address certain limitations. The proposed (CNN) model, while achieving remarkable performance, comes with the requirement of an additional preprocessing step. It assumes that the video footage capturing the exercise is already cropped, precisely marking the start and end of the movement. This assumption implies a reliance on preprocessed data, potentially limiting the model's adaptability to scenarios where such precise cropping is not readily available. Consequently, the generalizability of the proposed model may be influenced by the availability and accuracy of preprocessed datasets, necessitating careful consideration and potentially constraining its applicability in real-world situations where detailed annotations may be challenging to obtain.

8 Conclusion and future work

In conclusion, this research has significantly advanced the field of physical therapy and exercise classification through innovative methodologies. The introduction of a pioneering approach to feature engineering, representing exercises as 1D vectors using various statistical techniques, offers a novel perspective for exercise representation. Additionally, optimizing model performance by applying a random search methodology for designing the architectural framework of the proposed models contributes to increased efficiency and a deeper understanding of model behavior.

A substantial contribution is evident in the detailed comparative analysis across four distinct models (LSTM, Bi-LSTM, CNN, and CNN-LSTM) using two diverse datasets. This comparative study provides valuable insights into the strengths and weaknesses of each model in exercise classification, offering a nuanced understanding of performance variations. Moreover, the study's experimentation with multiple datasets, utilizing cross-validation techniques

to derive results from testing data, ensures robust model generalization performance across diverse datasets, enhancing the credibility of the study's findings.

The CNN model exhibited outstanding accuracy, attaining scores of 93.08% and 99.7% on the KIMORE and UI-PRMD datasets, respectively. This surpasses the state-of-the-art on both datasets by 0.75 and 0.1%, respectively. Moreover, the model demonstrated notable proficiency in disease classification, enabling the detection of correct and incorrect exercise techniques and achieving a disease diagnosis accuracy of 89.87%.

Notably, the (CNN) model boasts a lightweight architecture beyond its exceptional performance, making it suitable for deployment on low-power devices such as mobile apps. Furthermore, its efficient use of resources contributes to overall computational efficiency over video-based and image-based approaches. The model exhibits strong generalization capabilities across diverse datasets, ensuring its applicability to varied scenarios. This amalgamation of superior performance, resource efficiency, and generalization proficiency positions the (CNN) model as a promising and versatile solution in the domain of exercise classification and disease identification within the field of physical rehabilitation.

Adopting MLOps methods contributed to the smooth integration of machine learning into the software development life-cycle, improving collaboration, reproducibility, and scalability. This approach ensures deployed models' long-term efficacy and durability in dynamic operational situations.

In summary, this research's multifaceted contributions not only advance the understanding of physical rehabilitation and exercise classification but also extend the utility of the proposed models to disease identification. This marks a significant advancement in research and practical applications, contributing to the broader knowledge base and fostering future developments in the field.

Moreover, an area warranting further investigation is the accuracy of the KIMORE dataset. Despite achieving exceptional results, there is room for in-depth study to uncover potential refinements and optimizations. This exploration could delve into refining the model architecture, feature engineering techniques, or considering additional data augmentation strategies. Future research endeavors could also explore attention-based mechanisms to enhance model robustness. Specifically, implementing a fixed or overlapping window strategy might overcome the precision limitations associated with requiring a precise crop in video data, facilitating real-time applications. This approach could improve the model's adaptability to scenarios with challenging detailed annotations or precise cropping. Additionally, integrating Transfer Learning into our research paradigm is a promising avenue, albeit challenging. The dissimilarities between the UI-PRMD and KIMORE datasets, encompassing variations in anatomical joint data and exercise types, present hurdles that need creative solutions. Addressing these challenges will pave the way for more comprehensive and effective model transferability, ensuring the model's success in diverse settings and scenarios.

Author Contributions M. Zaher was responsible for implementing the methodology, conducting data analysis and visualization, and drafting the manuscript. A. Atia led the project, contributing to experimental design, interpreting findings, conceptualizing the study, and participating in manuscript writing. L. Abdelhamid co-led the project, played a role in dataset selection, contributed to study conceptualization, and reviewed the initial manuscript draft. A. Ghoniem, L. Abdelhamid, and A. Atia offered guidance throughout the research process. All authors reviewed and approved the final manuscript. This manuscript has not been previously published and is not under consideration for publication elsewhere.

Availability of data and materials The datasets employed and examined in this study are the University of Idaho Physical Rehabilitation Movements Dataset (UI-PRMD) [42] and the Kinematic Assessment of Movement and Clinical Scores for Remote Monitoring of Physical Rehabilitation (KIMORE) dataset [43], authored by Marianna Capecci et al., at the University of Pisa, Italy. These datasets are publicly accessible.

Declarations

Ethics approval This is not relevant because this study didn't use any human participants.

Consent to participate Therefore, participation was not subject to ethical review or consent.

Consent for publication This is not relevant because this study didn't use any human participants. As a result, publication did not require consent.

Competing interests The authors declare they have no potential competing interests in this article's research, authorship, or publication.

References

1. Organization WH (2024) Rehabilitation. Available online. https://www.who.int/health-topics/rehabilitation#tab=tab_1. Accessed 1 Jan 2024
2. Cunha B, Ferreira R, Sousa AS (2023) Home-based rehabilitation of the shoulder using auxiliary systems and artificial intelligence: an overview. Sensors 23(16):7100. https://doi.org/10.3390/s23167100
3. Arntz A, Weber F, Handgraaf M, Lällä K, Korniloff K, Murtonen K-P, Chichaeva J, Kidritsch A, Heller M, Sakellari E et al (2023) Technologies in home-based digital rehabilitation: scoping review. JMIR rehabilitation and assistive technologies 10:43615. https://doi.org/10.2196/43615
4. Loellgen H, Zupet P, Bachl N, Debruyne A (2020) Physical activity, exercise prescription for health and home-based rehabilitation. Sustainability 12(24):10230. https://doi.org/10.3390/su122410230
5. Paravlic AH (2022) Motor imagery and action observation as appropriate strategies for home-based rehabilitation: a mini-review focusing on improving physical function in orthopedic patients. Front Psychol 13:826476. https://doi.org/10.3389/fpsyg.2022.826476
6. Correia FD, Molinos M, Luís S, Carvalho D, Carvalho C, Costa P, Seabra R, Francisco G, Bento V, Lains J (2022) Digitally assisted versus conventional home-based rehabilitation after arthroscopic rotator cuff repair: a randomized controlled trial. Amer J Phys Med & Rehab 101(3):237. https://doi.org/10.1097/PHM.0000000000001780
7. Fascio E, Vitale JA, Sirtori P, Peretti G, Banfi G, Mangiavini L (2022) Early virtual-reality-based home rehabilitation after total hip arthroplasty: a randomized controlled trial. J Clin Med 11(7):1766. https://doi.org/10.3390/jcm11071766
8. Correia F, Nogueira A, Magalhães I, Guimarães J, Moreira M, Barradas I, Teixeira L, Tulha J, Seabra R, Lains J, Bento V (2018) Home-based rehabilitation with a novel digital biofeedback system versus conventional in-person rehabilitation after total knee replacement: a feasibility study. Sci Rep 8(1):11299. https://doi.org/10.1038/s41598-018-29668-0
9. Bhola G, Vishwakarma D (2023) A review of vision-based indoor har: state-of-the-art, challenges, and future prospects. Multimed Tools Appl pp 1–41. https://doi.org/10.1007/s11042-023-15443-5

10. Debnath B, O'brien M, Yamaguchi M, Behera A (2022) A review of computer vision-based approaches for physical rehabilitation and assessment. Multimed Syst 28(1):209–239. https://doi.org/10.1007/s00530-021-00815-4

11. Davids J, Lidströmer N, Ashrafian H (2022) In: Lidströmer N, Ashrafian H (eds.) Artif Intell Physiother Rehabil pp 1789–1807. Springer, Cham. https://doi.org/10.1007/978-3-030-64573-1_339

12. Coronato A, Naeem M, De Pietro G, Paragliola G (2020) Reinforcement learning for intelligent healthcare applications: A survey. Artif Intell Med 109:101964. https://doi.org/10.1016/j.artmed.2020.101964

13. Silva PP, Rocha WF, Mazzoni LE, Andrade RM, Bento A, Rampinelli M, Almonfrey D (2023) A diagnostic room for lower limb amputee based on virtual reality and an intelligent space. Artif Intell Med p 102612. https://doi.org/10.1016/j.artmed.2023.102612

14. Peng Y (2022) Human skeleton-based transfer learning on personalized rehabilitation training recognition. PhD thesis, University Name. https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-485128

15. Qiu Y, Wang J, Jin Z, Chen H, Zhang M, Guo L (2022) Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomed Signal Process Control 72:103323. https://doi.org/10.1016/j.bspc.2021.103323

16. Chen Y, Zhong K, Zhang J, Sun Q, Zhao X (2016) Lstm networks for mobile human activity recognition. In: 2016 International conference on artificial intelligence: technologies and applications, pp 50–53. https://doi.org/10.2991/icaita-16.2016.13. Atlantis Press

17. Mekruksavanich S, Jitpattanakul A (2021) Lstm networks using smartphone data for sensor-based human activity recognition in smart homes. Sensors 21(5):1636. https://doi.org/10.3390/s21051636

18. Domingo JD, Gómez-García-Bermejo J, Zalama E (2022) Improving human activity recognition integrating lstm with different data sources: Features, object detection and skeleton tracking. IEEE Access 10:68213–68230. https://doi.org/10.1109/ACCESS.2022.3186465

19. Wang C, Olugbade TA, Mathur A, De CWAC, Lane ND, Bianchi-Berthouze N (2019) Recurrent network based automatic detection of chronic pain protective behavior using mocap and semg data. In: Proceedings of the 2019 ACM international symposium on wearable computers, pp 225–230. https://doi.org/10.1145/3341163.3347728

20. Su J, Liao Z, Sheng Z, Liu A-X, Singh D, Lee H-N (2022) Human activity recognition using self-powered sensors based on multilayer bi-directional long short-term memory networks. IEEE Sens J. https://doi.org/10.1109/JSEN.2022.3195274

21. He J, Wu X, Cheng Z, Yuan Z, Jiang Y (2021) Db-lstm: Densely-connected bi-directional lstm for human action recognition. Neurocomputing 444:319–331. https://doi.org/10.1016/j.neucom.2020.05.118

22. Arif S, Wang J (2021) Bidirectional lstm with saliency-aware 3d-cnn features for human action recognition. J Eng Res 9(3A). https://doi.org/10.36909/jer.v9i3A.8383

23. Lee S-M, Yoon SM, Cho H (2017) Human activity recognition from accelerometer data using convolutional neural network. In: 2017 IEEE International conference on big data and smart computing (BigComp), pp 131–134. https://doi.org/10.1109/BIGCOMP.2017.7881728

24. Zeng M, Nguyen LT, Yu B, Mengshoel OJ, Zhu J, Wu P, Zhang J (2014) Convolutional neural networks for human activity recognition using mobile sensors. In: 6th International conference on mobile computing, applications and services, pp 197–205. https://doi.org/10.4108/icst.mobicase.2014.257786

25. Ji S, Xu W, Yang M, Yu K (2013) 3d convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell 35(1):221–231. https://doi.org/10.1109/TPAMI.2012.59

26. Nouriani A, McGovern RA, Rajamani R (2022) Deep-learning-based human activity recognition using wearable sensors. IFAC-PapersOnLine 55(37):1–6. https://doi.org/10.1016/j.ifacol.2022.11.152

27. Yadav SK, Tiwari K, Pandey HM, Akbar SA (2022) Skeleton-based human activity recognition using convlstm and guided feature learning. Soft Comput pp 1–4 https://doi.org/10.1007/s00500-021-06238-7

28. Boukhennoufa I, Zhai X, McDonald-Maier KD, Utti V, Jackson J (2021) Improving the activity recognition using gmaf and transfer learning in post-stroke rehabilitation assessment. In: 2021 IEEE 19th World symposium on applied machine intelligence and informatics (SAMI), pp 000391–000398. https://doi.org/10.1109/SAMI50585.2021.9378670

29. Réby K, Dulau I, Dubrasquet G, Aimar MB (2023) Graph transformer for physical rehabilitation evaluation. In: 2023 IEEE 17th International conference on automatic face and gesture recognition (FG), IEEE, pp 1–8. https://doi.org/10.1109/FG57933.2023.10042778

30. Mutlag AA, Abd Ghani MK, Mohd O, Abdulkareem KH, Mohammed MA, Alharbi M, Al-Araji ZJ (2023) A new fog computing resource management (frm) model based on hybrid load balancing and scheduling for critical healthcare applications. Phys Commun 59:102109. https://doi.org/10.1016/j.phycom.2023.102109

31. Hu H, Dong S, Zhao Y, Lian D, Li Z, Gao S (2022) Transrac: Encoding multi-scale temporal correlation with transformers for repetitive action counting. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 19013–19022. https://doi.org/10.48550/arXiv.2204.01018

32. Zhang Y, Shao L, Snoek CG (2021) Repetitive activity counting by sight and sound. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 14070–14079. https://doi.org/10.48550/arXiv.2103.13096

33. Dwibedi D, Aytar Y, Tompson J, Sermanet P, Zisserman A (2020) Counting out time: Class agnostic video repetition counting in the wild. In: Proceedings of the IEEE/cvf conference on computer vision and pattern recognition, pp 10387–10396. https://doi.org/10.48550/arXiv.2006.15418

34. Liao Y, Vakanski A, Xian M (2020) A deep learning framework for assessing physical rehabilitation exercises. IEEE Trans Neural Syst Rehabil Eng 28(2):468–477. https://doi.org/10.1109/TNSRE.2020.2966249

35. Abedi A, Bisht P, Chatterjee R, Agrawal R, Sharma V, Jayagopi DB, Khan SS (2023) Rehabilitation exercise repetition segmentation and counting using skeletal body joints. https://doi.org/10.48550/arXiv.2304.09735

36. Akkem Y, Biswas SK, Varanasi A (2023) Smart farming using artificial intelligence: A review. Eng Appl Artif Intell 120:105899. https://doi.org/10.1016/j.engappai.2023.105899

37. Li C, Ling X, Xia S (2023) A graph convolutional siamese network for the assessment and recognition of physical rehabilitation exercises. In: Iliadis L, Papaleonidas A, Angelov P, Jayne C (eds.) Artificial neural networks and machine learning – ICANN 2023, pp 229–240. Springer, Cham. https://doi.org/10.1007/978-3-031-44216-2_19

38. Kumar P, Suresh S (2023) Deep-har: an ensemble deep learning model for recognizing the simple, complex, and heterogeneous human activities. Multimed Tools Appl pp 1–28. https://doi.org/10.1007/s11042-023-14492-0

39. Bruce A, Chen Y, Liu Z (2022) Egcn: An ensemble-based learning framework for exploring effective skeleton-based rehabilitation exercise assessment. In: Proceedings of the international joint conference on artificial intelligence (IJCAI), pp 511–517. https://doi.org/10.24963/ijcai.2022/511

40. Rahevar M, Ganatra A, Saba T, Rehman A, Bahaj SA (2023) Spatial-temporal dynamic graph attention network for skeleton-based action recognition. IEEE Access 11:21546–21553. https://doi.org/10.1109/ACCESS.2023.3247820

41. Barzegar Khanghah A, Fernie G, Roshan Fekr A (2023) Design and validation of vision-based exercise biofeedback for tele-rehabilitation. Sensors 23:1206. https://doi.org/10.3390/s23031206

42. Vakanski A, Jun H-P, Paul D, Baker R (2018) A data set of human body movements for physical rehabilitation exercises. Data (Basel) 3(1). https://doi.org/10.3390/data3010002

43. Capecci M, Ceravolo MG, Ferracuti F, Iarlori S, Monteriú A, Romeo L, Verdini F (2019) The kimore dataset: Kinematic assessment of movement and clinical scores for remote monitoring of physical rehabilitation. IEEE Transactions on neural systems and rehabilitation engineering 27(7):1436–1448. https://doi.org/10.1109/TNSRE.2019.2923060

44. Kwapisz JR, Weiss GM, Moore SA (2011) Activity recognition using cell phone accelerometers. ACM SIGKDD Explorations Newsl 12:74–82. https://doi.org/10.1145/1964897.1964918

45. Reiss A, Stricker D (2012) Introducing a new benchmarked dataset for activity monitoring. In: 2012 16th International symposium on wearable computers, pp 108–109. https://doi.org/10.1109/ISWC.2012.13

46. Nahid A-A, Sikder N, Rafi I (2020) Ku-har: An open dataset for human activity recognition. Pattern Recogn Lett 3:46–54. https://doi.org/10.1016/j.patrec.2021.02.024

47. Reyes-Ortiz J-L, Oneto L, Samá A, Parra X, Anguita D (2016) Transition-aware human activity ecognition using smartphones. Neurocomputing 171:754–767. https://doi.org/10.1016/j.neucom.2015.07.085

48. Tasnim N, Baek J-H (2023) Dynamic edge convolutional neural network for skeleton-based human action recognition. Sensors 23(2):778. https://doi.org/10.3390/s23020778

49. Chen C, Jafari R, Kehtarnavaz N (2015) Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In: 2015 IEEE International conference on image processing (ICIP), pp 168–172. https://doi.org/10.1109/ICIP.2015.7350781

50. Li W, Zhang Z, Liu Z (2010) Action recognition based on a bag of 3d points. In: 2010 IEEE Computer society conference on computer vision and pattern recognition - workshops, pp 9–14. https://doi.org/10.1109/CVPRW.2010.5543273

51. Wang L, Zhang X, Zhang Y, Wang L (2019) A hybrid lstm-cnn model for skeleton-based action recognition. IEEE Transactions on Cybernetics 49(7):2608–2621. https://doi.org/10.1007/s00500-021-06238-7

52. Chen J, Wang J, Yuan Q, Yang Z (2023) Cnn-lstm model for recognizing video-recorded actions performed in a traditional chinese exercise. IEEE Journal of Translational Engineering in Health and Medicine. https://doi.org/10.1109/JTEHM.2023.3282245

53. Li H, Guo S, Bu D, Wang H, Kawanishi M (2023) Subject-independent estimation of continuous movements using cnn-lstm for a home-based upper limb rehabilitation system. IEEE Robotics and Automation Letters. https://doi.org/10.1109/LRA.2023.3303701

54. Lakhan A, Mohammed MA, Abdulkareem KH, Hamouda H, Alyahya S (2023) Autism spectrum disorder detection framework for children based on federated learning integrated cnn-lstm. Comput Biol Med 166:107539. https://doi.org/10.1016/j.compbiomed.2023.107539

55. Dong X, Yu Z, Cao W, Shi Y, Ma Q (2020) A survey on ensemble learning. Front Comp Sci 14:241–258. https://doi.org/10.1007/s11704-019-8208-z

56. Li C, Cheng L, Yang H, Zou Y, Huang F (2020) An automatic rehabilitation assessment system for hand function based on leap motion and ensemble learning. Cybern Syst 52(1):3–25. https://doi.org/10.1080/01969722.2020.1827798

57. Kato C, Uemura O, Sato Y, Tsuji T (2023) Functional outcome prediction after spinal cord injury using ensemble machine learning. Arch Phys Med Rehabil. https://doi.org/10.1016/j.apmr.2023.08.011

58. Wei W, McElroy C, Dey S (2019) Towards on-demand virtual physical therapist: Machine learning-based patient action understanding, assessment and task recommendation. IEEE Trans Neural Syst Rehabil Eng 27(9):1824–1835. https://doi.org/10.1109/TNSRE.2019.2934097

59. Yan S, Xiong Y, Lin D (2018) Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Thirty-second AAAI conference on artificial intelligence. https://doi.org/10.1609/aaai.v32i1.12328

60. Wang Z, Oates T et al (2015) Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. In: Workshops at the twenty-ninth AAAI conference on artificial intelligence, AAAI Menlo Park, CA, USA, vol 1. https://doi.org/10.48550/arXiv.1409.1556

61. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. https://doi.org/10.48550/arXiv.1409.1556

62. Cao Z, Hidalgo G, Simon T, Wei S-E, Sheikh Y (2021) OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans Pattern Anal Mach Intell 43(1):172–186. https://doi.org/10.48550/arXiv.1611.08050

63. Bazarevsky V, Grishchenko I, Raveendran K, Zhu T, Zhang F, Grundmann M (2020) BlazePose: On-device real-time body pose tracking. arXiv:2006.10204 [cs.CV]

64. Mroz S, Baddour N, McGuirk C, Juneau P, Tu A, Cheung K, Lemaire E (2021) Comparing the quality of human pose estimation with blazepose or openpose. In: 2021 4th International conference on bio-engineering for smart technologies (BioSMART), pp 1–4. https://doi.org/10.1109/BioSMART54244.2021.9677850

65. Lugaresi C, Tang J, Nash H, McClanahan C, Uboweja E, Hays M, Zhang F, Chang C-L, Yong MG, Lee J, Chang W-T, Hua W, Georg M, Grundmann M (2019) MediaPipe: A framework for building perception pipelines. arXiv:1906.08172 [cs.DC]

66. Slim SO, Atia A, Marwa M, Mostafa M- (2019) Survey on human activity recognition based on acceleration data. Int J Adv Comput Sci Appl 10(3). https://doi.org/10.14569/IJACSA.2019.0100311

67. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) Smote: Synthetic minority over-sampling technique. J Artif Intell Res 16:321–357. https://doi.org/10.1613/jair.953

68. Ronao CA, Cho S-B (2015) Deep convolutional neural networks for human activity recognition with smartphone sensors. In: Neural information processing: 22nd international conference, ICONIP 2015, Proceedings, Part IV 22, Springer, pp 46–53. Accessed 9–12 Nov 2015

69. Hartmann Y, Liu H, Schultz T (2022) High-level features for human activity recognition and modeling. In: International joint conference on biomedical engineering systems and technologies, Springer, pp 141–163. https://doi.org/10.1007/978-3-031-38854-5_8

70. Trelinski J, Kwolek B (2019) Ensemble of classifiers using cnn and hand-crafted features for depth-based action recognition. In: artificial intelligence and soft computing: 18th international conference, ICAISC 2019, Zakopane, Poland, Proceedings, Part II 18, Springer, pp 91–103. https://doi.org/10.1007/978-3-030-20915-5_9. Accessed 16–20 June 2019

71. Lemieux N, Noumeir R (2020) A hierarchical learning approach for human action recognition. Sensors 20(17):4946. https://doi.org/10.3390/s20174946

72. Yu T, Zhu H (2020) Hyper-parameter optimization: A review of algorithms and applications. https://doi.org/10.48550/arXiv.2003.05689

73. Feurer M, Hutter F (2019) In: Hutter F, Kotthoff L, Vanschoren J (eds.) Hyperparameter Optimization, pp 3–33. Springer, Cham. https://doi.org/10.1007/978-3-030-05318-5_1

74. Bergstra J, Bengio Y (2012) Random search for hyperparameter optimization. J Mach Learn Res 13:281–305

75. Zaremba W, Sutskever I, Vinyals O (2014) Recurrent neural network regularization. https://doi.org/10.48550/arXiv.1409.2329

76. Wang J, Sun S, Sun Y (2021) A muscle fatigue classification model based on lstm and improved wavelet packet threshold. Sensors 21(19):6369. https://doi.org/10.3390/s21196369

77. Huo Y, Wong DF, Ni LM, Chao LS, Zhang J (2020) Knowledge modeling via contextualized representations for lstm-based personalized exercise recommendation. Inf Sci 523:266–278. https://doi.org/10.1016/j.ins.2020.03.014

78. Wang Y, Zhang C, Zhao Y, Liao Y, Gao Y, Zheng J (2023) A method of classification decision based on multi-bilstms for physical loads hierarchy. Comput Methods Biomech Biomed Engin 26(10):1101–1113. https://doi.org/10.1080/10255842.2022.2106785

79. Hao Z, Kang Y, Dang X, An Y et al (2022) Wi-exercise: An indoor human movement detection method based on bidirectional lstm attention. Mobile Inform Syst 2022. https://doi.org/10.1155/2022/9994304

80. Hasan MW, Ferdosh Nima J, Sultana N, Ahmed Foysal MF, Karim E (2021) Exer-nn: Cnn-based human exercise pose classification. In: Emerging technologies in data mining and information security: proceedings of IEMIS 2020, vol 2, Springer, pp 347–361. https://doi.org/10.1007/978-981-13-1951-8

81. Mekruksavanich S, Jitpattanakul A (2022) Cnn-based deep learning network for human activity recognition during physical exercise from accelerometer and photoplethysmographic sensors. In: Pandian AP, Fernando X, Haoxiang W (eds.) Computer Networks, Big Data and IoT, pp 531–542. Springer, Singapore. https://doi.org/10.1007/978-981-19-0898-9_42

82. Arrowsmith C, Burns D, Mak T, Hardisty M, Whyne C (2022) Physiotherapy exercise classification with single-camera pose detection and machine learning. Sensors 23(1):363. https://doi.org/10.3390/s23010363

83. Vijayvargiya A, Khimraj Kumar R, Dey N (2021) Voting-based 1d cnn model for human lower limb activity recognition using semg signal. Phys Eng Sci Med 44:1297–1309. https://doi.org/10.1007/s13246-021-01071-6

84. Napieralski JA, Tylman W, Kotas R, Marciniak P, Kamiński M, Janc M, Józefowicz-Korczyńska M, Zamysłowska-Szmytke E (2022) Classification of subjects with balance disorders using 1d-cnn and inertial sensors. IEEE Access 10, 127610–127619. https://doi.org/10.1109/ACCESS.2022.3225521

85. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

86. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958

87. Bridle JS (1989) Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. In: advances in neural information processing systems, vol 2

88. Schuster M, Paliwal KK (1997) Bidirectional recurrent neural networks. IEEE Trans Signal Process 45(11):2673–2681. https://doi.org/10.1109/78.650093

89. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324. https://doi.org/10.1109/5.726791

90. Quan C, Ren K, Luo Z (2021) A deep learning based method for parkinson's disease detection using dynamic features of speech. IEEE Access 9:10239–10252. https://doi.org/10.1109/ACCESS.2021.3051432

91. Al Rahhal MM, Bazi Y, Al Zuair M, Othman E, BenJdira B (2018) Convolutional neural networks for electrocardiogram classification. J Med Biol Eng 38:1014–1025. https://doi.org/10.1007/s40846-018-0389-7

92. Wang K, Jiang P, Meng J, Jiang X (2022) Attention-based densenet for pneumonia classification. IRBM 43(5):479–485. https://doi.org/10.1016/j.irbm.2021.12.004

93. Lauritsen SM, Kalør ME, Kongsgaard EL, Lauritsen KM, Jørgensen MJ, Lange J, Thiesson B (2020) Early detection of sepsis utilizing deep learning on electronic health record event sequences. Artif Intell Med 104:101820. https://doi.org/10.1016/j.artmed.2020.101820

94. Donahue J, Hendricks LA, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, Darrell T (2015) Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2625–2634

95. Akkem Y, Biswas SK, Varanasi A (2023) Smart farming monitoring using ml and mlops. In: international conference on innovative computing and communication, Springer, pp 665–675. https://doi.org/10.1007/978-981-99-3315-0_51

96. Yaganteeswarudu A (2020) Multi disease prediction model by using machine learning and flask api. In: 2020 5th International conference on communication and electronics systems (ICCES), pp 1242–1246. https://doi.org/10.1109/ICCES48766.2020.9137896

97. Yaganteeswarudu A, Dasari P (2021) Diabetes analysis and risk calculation – auto rebuild model by using flask api. In: Chen JI-Z, Tavares JMRS, Shakya S, Iliyasu AM (eds.) Image Processing and Capsule Networks, Springer, Cham. pp 299–308. https://doi.org/10.1007/978-3-030-51859-2_27

98. Prechelt L (1998) In: Orr GB, Müller K-R (eds.) Early Stopping – But When?, Springer, Berlin, Heidelberg, pp 55–69. https://doi.org/10.1007/3-540-49430-8_3
99. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. https://doi.org/10.48550/arXiv.1412.6980
100. Zaher M, Samir A, Ghoneim A, Abdelhamid L, Atia A (2023) A framework for assessing physical rehabilitation exercises. In: 2023 Intelligent methods, systems, and applications (IMSA), pp 526–532. https://doi.org/10.1109/IMSA58542.2023.10217392

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Moamen Zaher[1,5] (ID) **· Amr S. Ghoneim[2]** (ID) **· Laila Abdelhamid[3]** (ID) **· Ayman Atia[4,5]** (ID)

Amr S. Ghoneim
amr.ghoneim@fci.helwan.edu.eg

Laila Abdelhamid
Laila.abdelhamid@fci.helwan.edu.eg

Ayman Atia
aezzat@msa.edu.eg
https://www.fcih.net/ayman/

[1] Software Engineering Program, Faculty of Computers & Artificial Intelligence—Helwan University, Ain Helwan 11795, Cairo, Egypt

[2] Computer Science Department, Faculty of Computers & Artificial Intelligence—Helwan University, Ain Helwan 11795, Cairo, Egypt

[3] Information System Department, Faculty of Computers & Artificial Intelligence—Helwan University, Ain Helwan 11795, Cairo, Egypt

[4] HCI-LAB, Faculty of Computers & Artificial Intelligence—Helwan University (HU), Ain Helwan 11795, Cairo, Egypt

[5] Faculty of Computer Science, October University for Modern Sciences and Arts (MSA), October 12451, Giza, Egypt