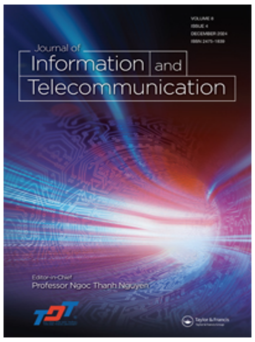

Journal of Information and Telecommunication

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tjit20

Rehabilitation monitoring and assessment: a comparative analysis of feature engineering and machine learning algorithms on the UI-PRMD and KIMORE benchmark datasets

Moamen Zaher, Amr S. Ghoneim, Laila Abdelhamid & Ayman Atia

JOURNAL OF INFORMATION AND TELECOMMUNICATION
https://doi.org/10.1080/24751839.2025.2454053

Rehabilitation monitoring and assessment: a comparative analysis of feature engineering and machine learning algorithms on the UI-PRMD and KIMORE benchmark datasets

Moamen Zaher [a,b], Amr S. Ghoneim [c], Laila Abdelhamid [d] and Ayman Atia [e,b]

[a]Software Engineering Prog. Faculty of Computing and Artificial Intelligence, Helwan University, Cairo, Egypt; [b]Faculty of Computer Science, October University for Modern Sciences and Arts (MSA), Giza, Egypt; [c]Computer Science Dept. Faculty of Computing and Artificial Intelligence, Helwan University, Cairo, Egypt; [d]Information System Dept. Faculty of Computing and Artificial Intelligence, Helwan University, Cairo, Egypt; [e]HCI-LAB, Faculty of Computing and Artificial Intelligence, Helwan University, Cairo, Egypt

ABSTRACT

Rehabilitation is crucial for individuals recovering from injuries or illnesses. It combines medical knowledge, therapy, and technology to improve health and independence. However, a global shortage of physiotherapists makes it challenging to provide adequate rehabilitation services. Current rehabilitation research often lacks advanced computational techniques to automate exercise assessment, relying heavily on time-consuming and costly in-person sessions. This study uses computer vision and classical machine learning (ML) to monitor and evaluate physical rehabilitation exercises using skeletal data. It compares five feature extraction approaches, six feature ranking techniques, and thirteen ML algorithms to identify the most influential features for accurate exercise classification using benchmark datasets (UI-PRMD and KIMORE). The performances of feature ranking algorithms–X2, ReliefF, Gini Decrease, FCBF, Information Gain, and Information Gain Ratio–were examined alongside ML algorithms such as SVMs, RFs, KNN, LDA, and lightGBM, amongst others. ReliefF with an Extra-Tree demonstrated superior performance (classification accuracy of 99.94%) compared to state-of-the-art studies on the UI-PRMD (a 4.4% improvement). However, FCBF, alongside an Extra-Tree, demonstrated robust performance across diverse datasets, achieving 99.64% on UIPRMD (the second-best result) and 81.85% on KIMORE (the highest accuracy reported compared to state-of-the-art studies). FCBF attained robust results together with the various classifiers, averaging 92.65%.

ARTICLE HISTORY
Received 18 September 2023
Accepted 11 January 2025

KEYWORDS
Exercise classification; Feature Ranking; Kinect; Physical Rehabilitation; Machine Learning

CONTACT Moamen Zaher moamenibrahim_psw@fci.helwan.edu.eg; moibrahem@msa.edu.eg Software Engineering Prog. Faculty of Computing and Artificial Intelligence, Helwan University, 11795, Cairo, Egypt; Faculty of Computer Science, October University for Modern Sciences and Arts (MSA), 12451, Giza, Egypt

1. Introduction

The field of Human Activity Recognition (HAR) has been extensively explored within the computer vision community. HAR encompasses various aspects, including Pose Estimation, Pose/Gesture Recognition, and Activity Detection. Evaluating human motion remains crucial, especially for applications like monitoring the functional rehabilitation of individuals with musculoskeletal disabilities (specifically affecting the muscles, bones, and joints) or physical disabilities in general. A significant body of research falls under categories such as Pose Estimation (S. Zhang et al., 2022), Human-Object Interaction (Im et al., 2018; Nagarajan et al., 2019), Activity Recognition (Chen et al., 2019; Herath et al., 2017; Patsadu et al., 2012; Poppe, 2010), Gesture Recognition (Lui, 2012), and Human-Human Interaction (Duarte et al., 2018; Sabale & Vaidya, 2016).

Rehabilitation, aimed at enhancing functioning and reducing disability in individuals, plays a vital role in healthcare (Zaher et al., 2025). Rehabilitation is essential for addressing a range of acute (typically short-lived/sudden) and chronic health conditions and complements medical and surgical interventions. It also helps manage complications associated with various health issues, such as fractures, strokes, and spinal cord injuries. Thus, it reduces hospitalization, shorter stays, and fewer readmissions. Many individuals worldwide, including those who have recovered from COVID-19, frequently need rehabilitation services to address ongoing health issues and improve their quality of life. Consequently, computer vision is utilized to accurately classify and monitor rehabilitation exercises, enhancing the effectiveness of physical therapy programmes. Several studies have explored physical rehabilitation exercise classification (Ahad et al., 2019; Debnath et al., 2022; Lei et al., 2019; Liao, Vakanski, Xian, Paul, et al., 2020).

In this paper, we present the extended and revised version of our research. Our proposed system aims to effectively alleviate the need for constant in-person supervision during exercises. This allows patients to safely perform rehabilitation routines at home, thereby reducing the frequency of clinic visits and lessening the workload of physical therapists. The system monitors and evaluates the patient's progress over time, ensuring effective rehabilitation. Each exercise is characterized by specific joint positions, angles, and trajectories. Ultimately, this system contributes to a reduction in rehabilitation costs.

In summary, the patient performs exercises in front of a Kinect camera. The system then extracts video frames, identifies body joints using Kinect, and classifies exercises to assess correctness. This process allows for accurate and real-time monitoring of the patient's rehabilitation progress, ensuring that they are performing the exercises correctly and safely.

The primary objective of our study is to assess the performance and effectiveness of various machine learning algorithms when paired with different feature selection techniques. This comparison is essential as it aids in identifying the most suitable methods for the problem discussed in this research. By systematically evaluating these techniques, our research offers valuable insights into their strengths and limitations. These findings can assist practitioners in selecting the appropriate methods for their applications and contribute to the ongoing development and refinement of these techniques.

The research covers three main areas: (1) Comprehensively comparing six different feature selection techniques, (2) Thoroughly evaluating thirteen machine learning algorithms for classifying physical rehabilitation techniques, and (3) Creating a heatmap to show the importance of body joints in classification.

2. Related work: approaches in rehabilitation exercise analysis

Researchers have explored various approaches in the field of exercise classification and evaluation, utilizing deep learning and innovative techniques: Rashid et al. (2020) investigated exercise style classification by incorporating spike train data into deep learning. The spike train data was chosen because it can capture distinct patterns in therapeutic movements. A deep convolutional network was employed for classification, achieving an accuracy score of 0.77%. Liao, Vakanski, and Xian (2020) also proposed a deep learning architecture for evaluating physical rehabilitation exercises. Their framework combined a convolutional neural network (CNN) for feature extraction from exercise videos with a long short-term memory (LSTM) network for activity classification. The method demonstrated strong performance, with average absolute deviations per exercise yielding a score of 0.02527.

Various deep learning architectures, including convolutional and recurrent neural networks, were examined by W. Zhang, Su, et al. (2020). The authors explored the application of deep learning in recognizing and evaluating rehabilitation exercises using smart sensors. The study reported impressive results, with deep learning models achieving recognition accuracy of up to 98.9% when trained on a specific dataset. Further exploration of datasets and advanced deep learning architectures was suggested for potential accuracy improvements.

Tang et al. (2022) introduced PointDet++, a novel framework for object detection based on human local features and a Transformer encoder. PointDet++ combines local features and a transformer encoder to detect objects in images. It outperformed existing state-of-the-art object detection methods on benchmark datasets, achieving an average precision (AP) of 54.2% on the COCO test-dev split and 46.1% on the Pascal VOC dataset. Additionally, PointDet++ significantly improved object detection speed, running up to 10 times faster than previous methods.

A view-invariant method for assessing human movements without relying on skeleton data was developed by Sardari et al. (2020). Their two-stage convolutional neural network, VI-Net, generated view-invariant trajectory descriptors from RGB images and processed them for all body joints. VI-Net exhibited strong performance with an average rank correlation of 0.65 on unfamiliar views and 0.66 on cross-subject scenarios when trained with only two views.

3. Methodology: the proposed framework

The data for this study is acquired using a Kinect camera, which extracts skeletal-body joints to serve as features. Feature engineering is performed by ranking and selecting the most influential features. Various machine-learning techniques are then applied for exercise classification. Figure 1 provides an overview of the proposed framework. Six feature ranking techniques were employed to determine the most suitable features for input into the model. Additionally, a comparative study was conducted between 13 different machine learning classifiers to identify the model with the highest accuracy.

The methodology section is divided into four subsections, each corresponding to a specific block within Figure 1. Subsection A covers the data acquisition phase. Subsection B discusses data processing for each dataset, visualized in Figure 1(b), followed by

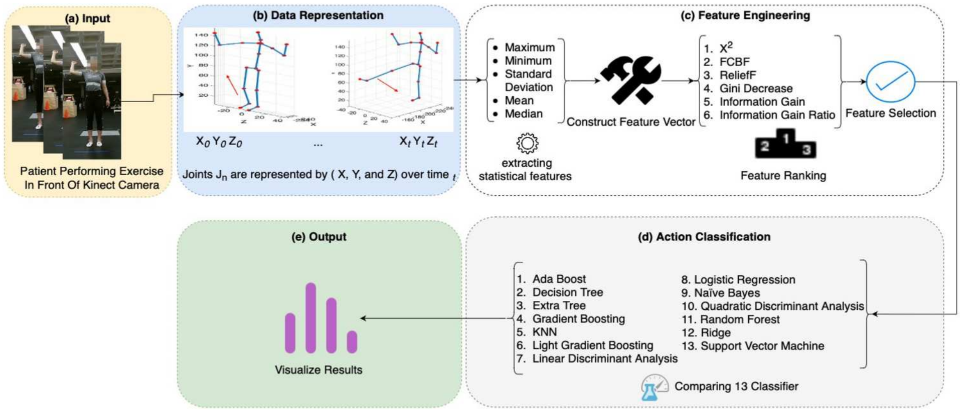

Figure 1. An overview of the proposed system.

detailing the feature selection techniques illustrated in Figure 1(c) in Subsection C. Finally, Subsection D describes the machine learning algorithms used for exercise classification, as shown in Figure 1(d).

3.1. Benchmark datasets: the UI-PRMD and KIMORE datasets

This research study conducted experiments using two distinct datasets: UI-PRMD (Liao, Vakanski, Xian, Paul, et al., 2020) and the KIMORE dataset (Capecci et al., 2019).

3.1.1. The UI-PRMD dataset

The users were required to perform the rehabilitation exercise with a Kinect camera. This camera processes videos at a rate of 30 frames per second and extracts the human skeleton from each frame at a maximum of 22 joints. The UI-PRMD dataset published by Liao, Vakanski, Xian, Paul et al. (2020) studied ten healthy individuals in a physical rehabilitation experiment. The dataset includes ten different exercises, with each individual asked to complete them correctly and incorrectly ten times while positioned in front of two motion-capturing systems: a Kinect camera and a Vicon optical tracker. The ten exercises studied were: Standing Shoulder Scaption, Hurdle Step, Standing Shoulder Extension, Side Lunge, Sit to Stand, Standing Shoulder Internal-External Rotation, Standing Active Straight Leg Raise, Standing Shoulder Abduction, Inline Lunge, and Deep Squat. Figure 2 shows examples of the 10 rehabilitation exercises. In this research, we focussed on the joints extracted from the Kinect Sensor which utilized the data gathered from the Kinect Camera.

This data was formatted in the YXZ coordinate system (Y: Height, X: Width, Z: Depth) for the following 22 body joints: Head, Head tip, Neck, Chest, Spine, Waist, Left Collar, Right Collar, Left Upper Arm, Right Upper Arm, Left Forearm, Right Forearm, Left Hand, Right Hand, Left Upper Leg, Right Upper Leg, Left Lower Leg, Right Lower Leg, Left Foot, Right Foot, Left Leg Toes, and Right Leg Toes. These joints and their angles were extracted from the sensors for each activity. Figure 3 illustrates a patient performing selected rehabilitation exercises in three-dimensional coordinates.

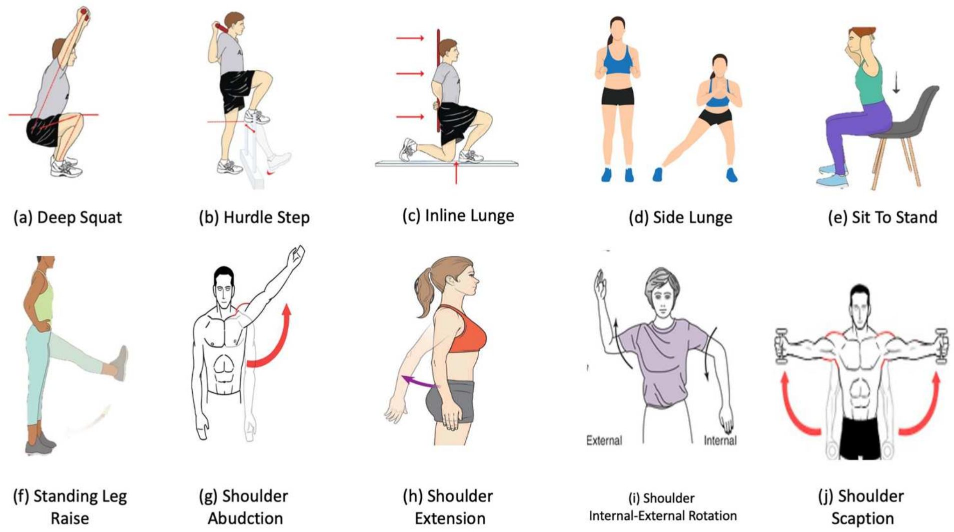

Figure 2. A graphical description of the ten physical rehabilitation exercises included in the UI-PRMD dataset.

As demonstrated in Figure 3, the waist area (a), feet and knees (b & c), and shoulder and neck (d) joints are the critical points for classifying their respective exercises. These observations are consistent across the 10 exercises and thus have been used to generate a heat map indicating the importance of each body joint for each exercise. Figure 1(a) depicts the patient performing the rehabilitation exercise before the Kinect camera, which captures the joint coordinates and uses them as system input. This dataset was a benchmark

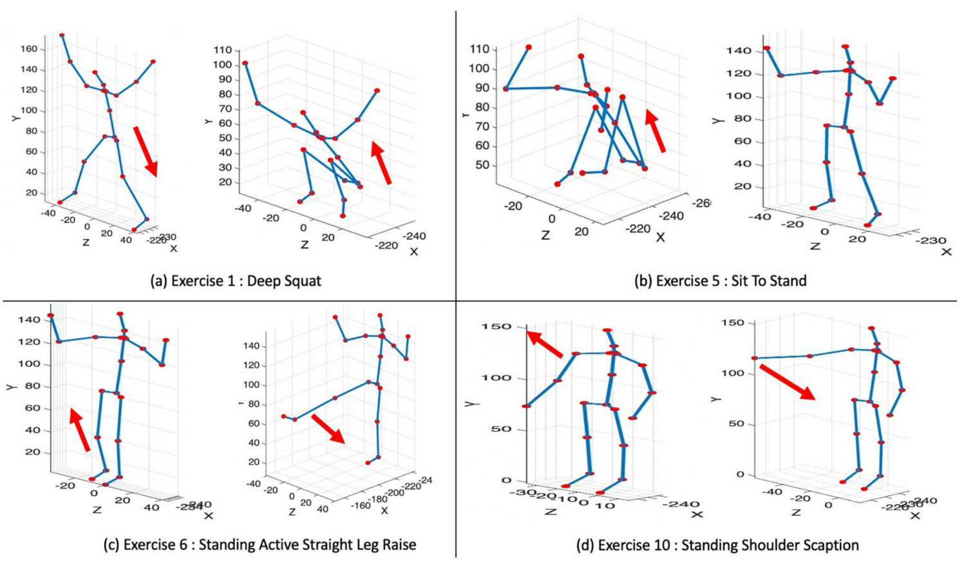

Figure 3. Three-dimensional imaging of a patient while they are undergoing selected rehabilitation exercises.

reference for training our model and conducting comparative analyses against previous research findings.

3.1.2. The KIMORE dataset

Kinematic Assessment of Movement and Clinical Scores for Remote Monitoring of Physical Rehabilitation (KIMORE) was authored by Capecci et al. (2019). It employs RGB-D sensors (Kinect v2) to record RGB and depth videos, along with skeletal joint positions, during five specific low back pain-focussed exercises. The Microsoft for Windows v2 sensor uses a novel Time- of-Flight (ToF) technology, while the previous sensor (Kinect v1) belongs to the category of Structured Light (SL) cameras. In contrast to SL cameras, ToF cameras exhibit an extended operational range and deliver depth maps with a higher degree of precision, devoid of any perceptible gaps or inaccuracies. The exercises include Lifting of Arms, Lateral tilt of the trunk with arms in extension, Trunk Rotation, Pelvic Rotation on the Transverse Plane, and Squatting. Data was captured using Kinect cameras in a controlled environment. Notably, the dataset comprises 78 subjects, including 44 healthy individuals and 34 with motor dysfunctions, providing diverse data for assessing computational approaches in rehabilitation scenarios. The dataset also provides a set of features for each exercise, defined explicitly by physicians, to describe its scope. These features, validated against a stereophotogrammetric system, can be analysed to compute a score for the subject's performance. The dataset comprises three subfolders: Raw data, Script, and Label. The participants are divided into two main groups: the Control Group (CG) and the Pain and Posture Disorders group (GPP). Within the CG, a further partition is observed, comprising two subgroups categorized as CG with expertize in physiotherapy exercises (CG-E) and CG without such expertize (CG-NE). Similarly, the GPP is segmented into three sub-groups based on specific diagnoses: Stroke, Parkinson's disease, and Low Back Pain. This research utilized the dataset for the purpose of disease classification rather than exercise categorization. The dataset was employed to illustrate the generalizability of the methodology to other datasets.

We recognize that Kinect technology, particularly its earlier versions, may not offer the highest level of accuracy compared to more modern solutions, such as mobile cameras with advanced APIs. However, the decision to use Kinect for this study was primarily based on the fact that both benchmark datasets, UI-PRMD and KIMORE, were built using Kinect sensors. These established datasets provide a solid foundation for our research, ensuring consistency, reproducibility, and comparability with existing studies. Despite the perception of Kinect's hardware as outdated, it remains a cornerstone in rehabilitation research due to its ability to capture skeletal data efficiently. Many techniques and methodologies in the literature still rely on Kinect-based data, providing a solid foundation for our analysis (Almasi et al., 2022; Ongun et al., 2020; Pedraza-Hueso et al., 2015). One of the key advantages of using skeletal data for classification is its flexibility in sensor selection, allowing the use of either Kinect or an RGB camera. By employing an API such as Mediapipe to extract joint information, this flexibility is maintained without compromising performance. A study by Zaher et al. demonstrated this adaptability by utilizing both Kinect and RGB data, applying the same model to skeletal data from both sources (Zaher, Samir, et al., 2023). Furthermore, using RGB cameras will cause privacy problems as sensitive information may be mistakenly obtained and exploited (Tasnim & Baek, 2023).

An alternative technology that may be considered is the use of thermal imaging, which could provide detailed information about temperature variations in the body during rehabilitation exercises. Incorporating thermal data would undoubtedly provide a new dimension of information, particularly concerning muscle activity or inflammation, but that would require different data sources and methodologies beyond the scope of our current analysis. The primary goal of our study is to assess the effectiveness of feature engineering and ML approaches using skeletal data, and Kinect's motion-capture capabilities remain sufficient for this purpose. Moreover, incorporating thermal imaging will increase the model complexity. This hinders its real-time applicability and will make deployment on resource-constrained edge devices more challenging (Tasnim & Baek, 2023).

3.2. Data pre-processing

Distinct preprocessing methods were implemented for each dataset in preparation for the feature engineering phase. Normalisation techniques were applied to all features to ensure that each feature's values were brought to a standard scale. This normalization step is essential to prevent features with larger scales from dominating those with smaller scales during the modelling process, allowing for fair and unbiased contributions from each feature. One-Hot Encoding was also employed to encode the labels of both datasets.

3.2.1. Preprocessing the UI-PRMD dataset

The Kinect Camera recorded 22 body joints, each represented as a three-dimensional vector Jn with coordinates Xt, $Y t$, and Zt over time t and stored in a vector V. Feature extraction techniques were applied to these joints, including statistical measures such as maximum, minimum, mean, standard deviation, and median. This process resulted in a total of 330 features, which were subsequently subjected to feature selection for model enhancement.

Figure 1(b) visualizes the data representation phase.

3.2.2. Preprocessing the KIMORE dataset

The original dataset comprises $XY Z$ coordinates for 25 distinct body joints alongside a confidence score. This score represents the Kinect sensor's confidence in the 3D joint position, ranging from 1 (indicating high confidence) to 0 (indicating low confidence). In total, the dataset encompasses 100 features. Various statistical techniques, including standard deviation, maximum, median, mean, and minimum, were applied to each of these 100 features, resulting in a total of 500 features. Due to the imbalanced nature of the KIMORE dataset, we employed the Synthetic Minority Over-sampling Technique (SMOTE) technique to rectify the data distribution. SMOTE (Chawla et al., 2002) is a data augmentation approach employed in machine learning that aims to rectify class imbalance issues by creating synthetic instances within the minority class through interpolation between existing data points. This technique effectively mitigates skewed class distributions, consequently enhancing model performance, mainly when dealing with substantial class imbalances.

3.3. Feature selection

Feature selection is a technique to identify and select a subset of relevant features for model construction. It is designed to reduce the number of input variables when creating a predictive model and to improve the accuracy and interpretability of the model by selecting only the most pertinent features. Action classification in physical rehabilitation exercises is the process of categorizing an action or movement into one or more categories based on specific criteria. These criteria often include the type of movement, range of motion, speed or intensity of the movement, coordination, timing of the body parts involved, and any other relevant factors. Action classification is utilized to evaluate and monitor a patient's progression in physical rehabilitation, as well as to customize a patient's exercise plan to their specific requirements (Miron & Grosan, 2021).

Feature selection is implemented using various methods, such as filter, wrapper, and embedded methods. Filter methods measure the importance of features with statistical measures, while wrapper methods employ predictive models to assess feature subsets. Embedded methods utilize an optimization process to choose relevant features. This research utilized six distinct filter methods (Relief, FCBF, Chi-Squared, Gini Decrease, Information Gain, and Information Gain Ratio) to identify the most important features. The feature engineering phase is illustrated in Figure 1(c).

While the UI-PRMD dataset generated 330 features and the KIMORE dataset generated 500 features during preprocessing, applying feature ranking was essential to identify the most influential features for classification. The six feature ranking techniques selected are widely recognized and have demonstrated effectiveness in capturing relevant data points while reducing dimensionality. The decision to focus on six feature ranking techniques was based on the need to balance complexity, computational efficiency, and interpretability. In this study, we aimed to compare diverse approaches while maintaining practical applicability in the context of rehabilitation monitoring.

Moreover, a heatmap was generated to discern the predominant body joints correlated with the rehabilitation exercises. As depicted in Figure 4, the primary body joints exhibit a notable concentration around the waist region, with secondary significance

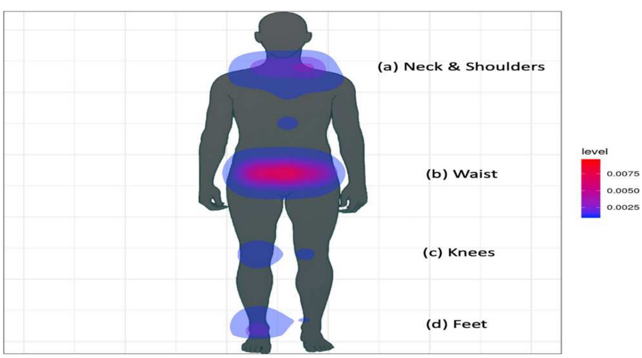

Figure 4. A heatmap of key joints to identify the key body joints; (a) the Neck and Shoulder Area, (b) the Waist Area, (c) the Knee Area, and (d) the Feet Area.

observed in the neck, shoulders, feet, and knees. This heatmap was constructed through a comprehensive evaluation of features derived from the six distinct feature ranking techniques.

During our experiments, we found that two feature ranking techniques, ReliefF and Fast Correlation Based Filter (FCBF), yielded the best results among the six techniques evaluated. Therefore, this subsection will focus on discussing these two methods in detail. This allows for a more in-depth analysis of the most effective techniques for identifying influential features in the context of exercise classification.

3.3.1. ReliefF

Relief algorithms are generally applied as a feature subset selection technique in the prepossessing phase before the model is constructed (Urbanowicz et al., 2018). Relief algorithms' core notion is to measure an attribute's quality by how effectively it distinguishes between instances that are close to one another. While many heuristic measures assume that the attributes are conditionally independent of one another to estimate the quality of the attributes, Relief methods do not. They are effective, cognizant of contextual information, and capable of accurately estimating the quality of features in circumstances when there are significant connections between the attributes. The original Relief is capable of dealing with both nominal and numerical properties. However, it cannot deal with incomplete data and is limited to two-class problems. The difference in the values of the attribute A for the two instances I_1 and I_2 is calculated by the function diff(A, I_1, I_2). For instance, the record of movement A is represented by the mean value of Waist X coordinates I_1, and the maximum value of Neck Y coordinates I_2. For numerical attributes, the equation is originally defined as in Equation (1).

$$diff(A, I_1, I_2) = \frac{|value(A, I_1) - value(A, I_2)|}{\max(A) - \min(A)} \quad (1)$$

ReliefF is the extension that solves these and other issues. The ReliefF technique was employed, consequently handling multi-class issues.

3.3.2. Fast correlation based filter (FCBF)

Fast Correlation Based Filter (FCBF) is an algorithm used for feature ranking and feature subset selection (Yu & Liu, 2004). It is a supervised feature ranking technique, which means that it considers the class label of the data to determine the relevance of the features. FCBF works by measuring the correlation between two features and the class label and using the correlation score to rank the features. The higher the correlation score, the more relevant the feature is to the class label. The FCBF algorithm is computationally efficient and is often used for feature selection in machine learning applications. Based on information theory, FCBF calculates feature dependencies and class relevance using symmetrical uncertainty. FCBF heuristically uses a backward selection method in conjunction with a sequential search approach to eliminate unnecessary and redundant features. When there are no more features to remove, the algorithm quits. The Fast Correlation-Based Filter (FCBF) algorithm consists of two steps: (1) Analysis of Relevance: the first step computes the correlation between each feature and the class label, (2) Redundancy Analysis: the second step ranks and filters the features based on their correlation scores. Analysis of Relevance. Correlation is frequently used to evaluate significance. The linear correlation coefficient is used to quantify correlation in linear systems. However, non-linear systems dominate in real-world applications. Symmetrical Uncertainty (*SU*) is used to measure correlation in non-linear systems as defined in Equations (2)–(4). The variables *X* and *Y* refer to the features of the dataset, such as spine and waist measurements.

$$SU = 2\left[\frac{IG(X \mid Y)}{H(X)H(Y)}\right] \tag{2}$$

$$IG(X, Y) = H(X) - H(X \mid Y) \tag{3}$$

$$H(X) = -\sum_{i} P(X_i) - \log_2 P(X_i) \tag{4}$$

Where $IG(X \mid Y)$ is the Information Gain of *X* after observing variable *Y*. $H(X)$ and $H(Y)$ are the entropy of variables *X* and *Y*, respectively. $P\,(xi)$ is the probability of variable *X*. *SU* is the modified version of Information Gain with a range between 0 and 1. FCBF removes irrelevant features by ranking correlation (*SU*) between feature and class. If *SU* between feature and class is equal to 1, it means that this feature is entirely related to that class. On the other hand, if *SU* is equal to 0, the features are irrelevant to this class.

$IG(X \mid Y)$ is the Information Gain of *X* following the observation of variable *Y*. The entropy of variables *X* and *Y* is represented by $H(X)$ and $H(Y)$, respectively. The probability of variable *X* is given by $P(xi)$. *SU* is a modified version of Information Gain with a range of 0 to 1. FCBF removes irrelevant features by ranking correlation (*SU*) between feature and class. If *SU* between feature and class is equal to 1, it means that this feature is entirely related to that class. If *SU* is equal to zero, the features are irrelevant to this class.

Redundancy Analysis. Upon ranking relevant features, FCBF eliminates redundant features from selected features based on *SU* between feature and class and between feature and feature. Redundant features are defined from the meaning of predominant feature and approximate Markov Blanket. In Yu and Liu (2004), a feature is predominant (both relevant and non-redundant feature) if it does not have an approximate Markov blanket in the current set. Approximate Markov blanket: for two relevant features F_i and F_j ($i \neq j$), F_j forms an approximate Markov blanket for F_i as demonstrated in Equation (5).

$$SU_{j,c} \geq SU_{i,c} \quad \text{and} \quad SU_{i,j} >= SU_{i,c} \tag{5}$$

The correlation between any feature and the class is $SU_{i,c}$. The correlation between any pair of features F_i and F_j is represented by $SU_{i,j}$. ($i \neq j$).

3.4. Action classification

Action classification models are divided into two groups: ensemble models, which utilize multiple models to improve the accuracy of predictions, and non-ensemble models, which make predictions using only one model.

In this paper, five ensemble models, namely Random Forest (RF), Ada Boost (ADA), Gradient Boosting Classifier (GBC), Extra Trees Classifier (ET), and Light Gradient Boosting Machine (lightGBM), were implemented. Additionally, eight non-ensemble models, such as Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), K Nearest Neighbors (KNN), Naive Bayes (NB), Ridge, Quadratic Discriminant Analysis (QDA), and

Linear Discriminant Analysis (LDA), were also utilized. Choosing which algorithm achieves the highest accuracy is the primary goal. Figure 1(d) clarifies the classification phase.

The decision to apply thirteen different algorithms, including both ensemble and non-ensemble models, was made to conduct a comprehensive evaluation of various machine-learning approaches for rehabilitation exercise classification. Given the complexity of the task and the diversity of the datasets (UI-PRMD and KIMORE), we sought to assess a wide range of algorithms to identify which techniques perform best under different conditions. By including a variety of algorithms, we aimed to capture the strengths of different methodologies. For instance, ensemble models such as Random Forest (RF) and Gradient Boosting (GBC) have been known to perform well with high-dimensional data (Lusa et al., 2017). In contrast, non-ensemble models like Support Vector Machines (SVM) and Logistic Regression (LR) offer interpretability and are computationally efficient (Huang et al., 2014). The goal was to provide a holistic comparison to guide future studies and practical applications in rehabilitation monitoring. While it may seem large, the inclusion of thirteen algorithms allows us to provide a well-rounded analysis and ensure that the selected models are truly the best fit for the data at hand.

During the experiments, we discovered that ensemble models consistently outperformed non-ensemble models. Consequently, this subsection focuses on discussing the three highest-achieving ensemble models rather than all thirteen algorithms tested. This choice ensures a more detailed analysis of the most effective approaches for exercise classification.

3.4.1. Extra tree classifier

Extra Tree Classifier is an ensemble tree-based technique for machine learning (Geurts et al., 2006). Which reduces variance and processing cost by using randomization (compared to Random Forest). Extra Tree develops decision trees using the complete dataset, as opposed to Random Forest which utilizes bagging to choose several variations of the training data to ensure that decision trees are sufficiently distinct. It randomly selects the values to split a feature and create child nodes. Compared to Random Forest, Extra Trees is substantially faster. The greedy approach employed in Random Forest is currently replaced by additional trees choosing at random the value at which to split features. This study implements an Extra Tree Classifier by limiting the number of samples required to be at a leaf node to one, the number of trees in the forest to 100 (to reduce variance), and the number of characteristics to be considered at each decision to one. Figure 5 Shows the architecture of the Extra Tree algorithm.

Palimkar et al. (2022) implemented an Extra Tree classifier for Human Activity Recognition (HAR) using a combination of magnetometer, accelerometer, and gyroscope data. The classifier could accurately identify activities, even in dynamic environments, showing its robustness across a wide range of scenarios.

3.4.2. Gradient boosting

Gradient Boosting (Natekin & Knoll, 2013) is a collection of decision trees that differs mostly in two ways: (a) Using gradient boosting, one tree is constructed at a time, whereas random forests build each tree independently, (b) Combining findings; Random forests average or use ‘majority rules’ to combine outcomes after the process, whereas gradient boosting

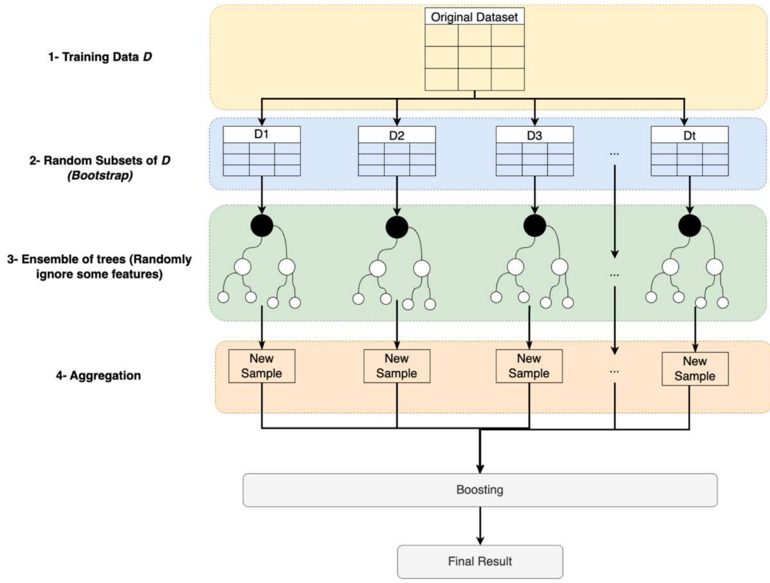

Figure 5. Extra tree classifier architecture.

combines results as it goes. Performance-wise, gradient boosting outperforms random forests. They are typically more difficult to tune than random forests.

3.4.3. LightGBM

A gradient-boosting framework called LightGBM (Machado et al., 2019) leverages tree-based learning methods. It is intended to be more effective than conventional gradient-boosting decision trees (GBDT) and is capable of processing huge amounts of data. Instead of expanding decision trees horizontally, branch by branch, as most other tree-based models do, LightGBM grows them vertically, level by level. Additionally, compared to typical GBMs, which examine many features at each level, training time is decreased because just one feature is taken into account at each level. LightGBM employs two innovative strategies, Gradient-Based One-Side Sampling, and Exclusive Feature Bundling, to address the drawbacks of the Histogram-Based Approach, which is widely used in all Gradient Boosting Decision Tree frameworks.

4. Experiments and results

The initial experiment aimed to compare state-of-the-art feature selection techniques with the goal of reducing the feature count and identifying the most effective feature ranking method. Subsequently, the second experiment involved the comparison of various classification algorithms using the same dataset to determine the optimal combination of feature ranking techniques and machine learning classification algorithms. Additionally, the third experiment sought to validate the model's generalizability. In this study, we assess the model's performance using accuracy as the evaluation metric, considering the balanced nature of the dataset (Sokolova & Japkowicz, 2006).

To ensure a fair comparison, all experiments were conducted on the same machine equipped with a Tesla T4 GPU with 15GB of GPU RAM and 12GB of system RAM. This configuration is common in many previous researches (Tamla et al., 2022; Yadav et al., 2023).

4.1. Experiment 1

Following feature extraction, a total of 330 features were present. Therefore, different feature ranking procedures were examined. The top 20 features from each of the six advanced feature ranking algorithms used in this study — $X2$, ReliefF, Gini Decrease, FCBF, Information Gain, and Information Gain Ratio — were then put through another experiment to ascertain the most efficient classification method.

The decision to limit the feature set to 20 features stems from evaluating the scores assigned by various ranking techniques to each feature. It was observed that beyond the 20th feature, the scores either significantly declined or reached a plateau. This pattern was consistent across all six feature ranking techniques. Figure 6 visually presents the scores assigned to the top 30 features by each of these ranking techniques.

ReliefF achieved the highest accuracy of 99.94%, followed by FCBF with an accuracy of 99.64%. Gini Decrease and Information Gain both had the best accuracy of 99.06%, tying for third place. The best accuracy for the Information Gain Ratio and X^2 techniques was 99.0%.

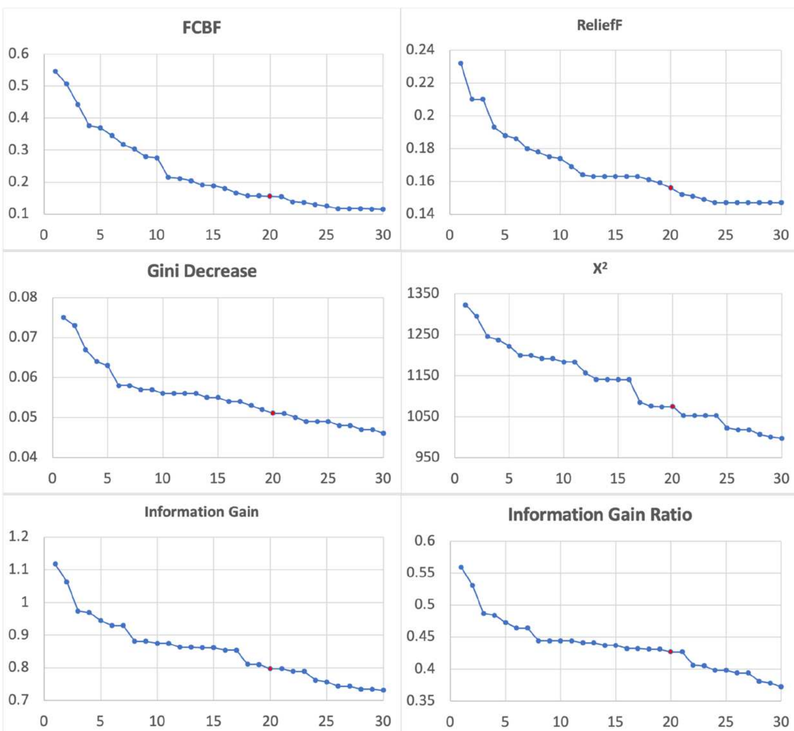

Figure 6. The importance score for each feature either plateaued or declined after 20 features.

Figure 7 shows the highest and average accuracy for each technique that has been recorded.

FCBF had the highest average accuracy score among all the classifiers tested of 92.65%, while ReliefF had the lowest average accuracy of 83.62%, despite achieving the highest recorded accuracy during the experiment. The X^2 technique was the runner-up with an average accuracy of 91.37%, followed by Gini Decrease and Information Gain Decrease, which attained an average accuracy of 88.95% and 87.28%, respectively. Following applying our comparison to the various classifiers, the average score of each technique is shown in Figure 7. Table 1 presents the outcomes obtained from the proposed method applied to the UI-PRMD dataset, evaluating 13 distinct models in conjunction with feature ranking techniques. The final column in Table 1 displays the average performance across all algorithms using the various feature ranking methods. This analysis is crucial as it provides insight into the sensitivity of each model to the chosen feature ranking approach. Ensemble models emerged with both superior performance and demonstrated lower sensitivity to the employed ranking techniques.

4.2. Experiment 2

In this experiment, the top 20 features were ranked, and the remaining features were excluded. The newly selected features were then fed into the model. The average score of each technique was determined by applying our comparison to the 13 different machine learning classifiers. The algorithms used were ADA, DT, ET, GBC, KNN, lightGBM, LDA, LR, NB, QDA, RF, Ridge and SVM. This research used 30% of the data for testing and 70% for training. Stratified splitting was employed to ensure an equitable distribution of classes both during training and testing. This approach helps maintain the same class distribution and mitigates the risk of biased training. Cross-validation of 30 folds was applied. Throughout the experiments, regardless of the feature ranking method employed, the Light Gradient Boosting Machine (LightGBM) consistently outperformed the other classifiers, with an average accuracy of 99.0%. Gradient Boosting (GBC) and Extra Tree (ET) consistently shared the second spot with average accuracy scores of 98.8% and 98.6%, respectively.

This research revealed that Gradient Boosting Machine (GBM) and Light Gradient Boosting Machine (LightGBM) needed more training time than the other classifiers, with LightGBM needing slightly less time than GBM. However, they took longer than

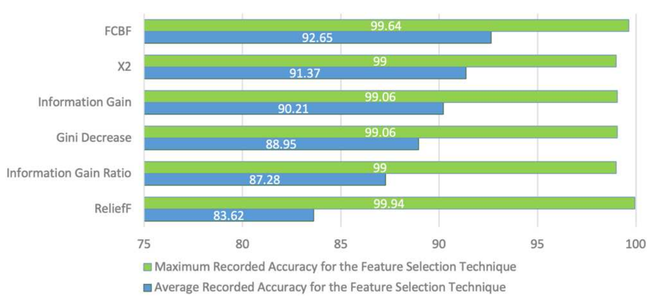

Figure 7. Illustration of the average and maximum recorded accuracy of different classifiers when feature ranking techniques are employed.

Table 1. The mean performance of various classifiers combined with different feature ranking techniques on UI-PRMD dataset after applying cross-validation.

Classifier	X^2	FCBF	ReliefF	Gini Decrease	p5pcInformation Gain	Information Gain Ratio	Mean Classifier Accuracy
Ada Boost Classifier	94.04%	93.37%	86.20%	93.66%	93.60%	91.66%	92.10%
Decision Tree	99.00%	99.07%	91.12%	98.81%	98.81%	98.81%	98.81%
Extra Tree	98.37%	99.64%	***99.94%***	97.81%	97.56%	98.25%	98.60%
Gradient Boosting Classifier	98.87%	98.93%	98.50%	98.81%	98.81%	98.81%	98.80%
KNN	92.44%	94.01%	99.88%	91.63%	91.63%	93.07%	93.80%
Light Gradient Boosting Machine	98.94%	99.36%	98.75%	99.06%	99.06%	99.00%	**99.00%**
Linear Discriminant Analysis	98.87%	68.05%	97.00%	97.62%	97.62%	97.56%	92.80%
Logistic Regression	90.75%	94.86%	79.12%	87.94%	87.94%	88.31%	88.20%
Naïve Bayes	96.94%	96.85%	95.25%	96.38%	96.38%	86.75%	94.80%
Quadratic Discriminant Analysis	66.18%	87.76%	18.44%	67.31%	68.25%	40.56%	58.10%
Random Forest	98.31%	99.57%	99.62%	82.24%	98.06%	98.06%	96.00%
Ridge	68.26%	77.71%	53.69%	59.81%	59.81%	59.94%	63.20%
Support Vector Machine	86.83%	95.25%	69.61%	85.25%	85.25%	83.82%	84.30 %

Extra Tree, Decision Tree, and Random Forest. When paired with the most effective technique, some algorithms achieved accuracy scores much higher than their average. For instance, a Support Vector Machine (SVM) had an average accuracy score of 84.3%, but when paired with a Fast Correlation-Based Filter (FCBF), its accuracy increased to 95.26%.

The Extra Tree showed remarkable results when combined with either the FCBF or the ReliefF feature ranking techniques. The Extra Tree Classifier combined with ReliefF had the highest accuracy score of 99.94%, followed by the KNN and the Random Forrest, with maximum accuracy scores of 99.88% and 99.62%, respectively. When combined with the FCBF feature ranking technique, the Decision Tree and Gradient Boosting Machine achieved maximum accuracy scores of 99.07% and 98.93%, respectively. Additionally, the Light Gradient Boosting Classifier had an accuracy score of 99.36%. Table 1 shows the achieved results of the proposed approach.

The proposed model demonstrates superior performance in comparison to prior studies conducted on the UI-PRMD dataset. Table 2 provides a comparative analysis of the outcomes achieved by previous research efforts and our best-performing model. Notably, our model exhibits an approximate 4.4% improvement over the previous work.

4.3. Experiment 3

This experiment was conducted to assess the generalizability of the methodology to diverse datasets. The outcomes of this experiment are displayed in Table 3. Previous studies have also examined the performance of various assessment techniques on both the KIMORE and UI-PRMD datasets. Notably, the UI-PRMD dataset is much easier to classify than the KIMORE dataset, as evidenced in earlier research (Réby et al., 2023).

In this dataset, the Information Gain feature ranking technique proved to be the most suitable choice in terms of mean score across algorithm. Furthermore, the Extra Tree algorithm outperformed all other machine learning algorithms. Specifically, when combined with the FCBF feature ranking technique, it achieved the highest accuracy of 81.85%. Additionally, Random Forest combined with the Information Gain Ratio feature ranking technique yielded an accuracy of 78.43%. LighGBM was third again with an accuracy of 78.43%.

5. Discussion

The tests revealed that the Extra Tree, KNN, and Random Forest algorithms are most effective when paired with the ReliefF feature ranking technique. In contrast, the Light Gradient Boosting Machine, Decision Tree, Gradient Boosting, Support Vector Machine, Logistic Regression, Quadratic Discriminant Analysis, and Ridge algorithms were found

Table 2. Comparing our proposed models to other related work on UI-PRMD dataset.

Algorithm	Accuracy	Precision	Recall
LSTM-CNN (Wang et al., 2019)	89.8%	90.7%	89.0%
3D-ResNet (Y. Zhang, Wang, et al., 2020)	91.7%	92.6%	90.8%
EGCN (Bruce et al., 2022)	93.3%	94.3%	92.3%
DeepPose (Niu et al., 2023)	95.1%	95.7%	94.5%
Graph Transformer for Physical Rehabilitation Evaluation (G2PRE) (Réby et al., 2023)	94.5%	94.1%	94.9%
ReliefF-Extra Tree	**99.94**%	**99.94**%	**99.94**%

Table 3. The mean performance of various classifiers combined with different feature ranking techniques on KIMORE dataset after applying cross-validation.

Classifier	X^2	FCBF	ReliefF	Gini Decrease	Information Gain	Information Gain Ratio	Mean Classifier Accuracy
Ada Boost	54.95%	56.67%	53.38%	55.28%	56.11%	59.86%	56.04%
Decision Tree	67.45%	74.35%	68.56%	66.53%	72.82%	73.19%	70.48%
Extra Tree	76.81%	***81.85***%	74.63%	72.78%	75.14%	76.81%	**76.34%**
Gradient Boosting Classifier	74.95%	74.95%	68.80%	69.54%	74.03%	73.98%	72.71%
KNN	75.00%	71.30%	71.85%	68.52%	71.90%	73.15%	71.95%
Light Gradient Boosting Machine	73.89%	71.44%	73.89%	72.96%	76.90%	77.64%	75.45%
Linear Discriminant Analysis	58.89%	58.89%	61.94%	61.53%	57.78%	60.83%	59.98%
Logistic Regression	59.12%	59.12%	53.80%	63.19%	58.52%	56.48%	58.37%
Naïve Bayes	57.92%	57.92%	54.44%	58.84%	59.26%	54.68%	57.18%
Quadratic Discriminant Analysis	18.19%	18.19%	32.45%	74.91%	75.56%	18.19%	39.58%
Random Forest	76.53%	76.53%	72.22%	70.46%	75.65%	78.43%	74.97%
Ridge	56.94%	56.94%	55.42%	56.30%	58.75%	57.92%	57.05%
Support Vector Machine	57.55%	66.20%	58.5%	58.06%	62.04%	58.43%	60.13 %

to be most successful when combined with the FCBF feature ranking technique. Additionally, the Naive Bayes, Logistic Regression, and Ada Boost Classifier algorithms produced the best results when used in combination with the $X2$ feature ranking technique. Regardless of the feature selection technique employed, this study showed that Light Gradient Boosting Classifier, Gradient Boosting, Extra Tree, and Random Forest consistently achieved the highest average scores on the UI-PRMD dataset.

For the KIMORE dataset, Extra Tree also attained the highest results, but this time in combination with FCBF. It was followed by Random Forest and LightGBM combined with the Information Gain Ratio. These findings highlight that Extra Tree and Random Forest consistently deliver the highest accuracy across all algorithms, irrespective of the dataset employed. Furthermore, the choice of feature ranking techniques significantly impacts the model's performance, with each ranking technique being more suitable for one dataset over another.

Although ReliefF demonstrated the highest top-1 accuracy on the UIPRMD dataset when paired with Extra Tree, it exhibited challenges when integrated with other models. Furthermore, ReliefF did not consistently produce reliable results across multiple benchmarks. In contrast, FCBF consistently delivered robust performance regardless of the classification algorithm used, as depicted by the average accuracy across various ranking techniques in Figure 7. Additionally, FCBF demonstrated strong performance on multiple benchmark datasets, as evidenced in Tables 1 and 3. Therefore, we recommend employing FCBF for feature ranking in conjunction with the ensemble model Extra Tree (ET) for classification purposes.

6. Conclusion

In conclusion, home-based rehabilitation presents a viable solution to address global shortages of physiotherapists and the limitations in funding and infrastructure for traditional rehabilitation settings. It enhances the productivity of each physiotherapist through remote patient monitoring. However, a significant challenge remains in ensuring the correct execution of exercises, as improper techniques can hinder recovery and prolong rehabilitation times. Therefore, accurate classification and assessment of exercise execution are crucial.

This research underscores the importance of methodological choices in enhancing the reliability and applicability of machine learning models for assessing exercise execution in home-based rehabilitation contexts. The study comprehensively evaluated 13 machine-learning methods for identifying 10 distinct rehabilitation exercises, employing various feature ranking techniques to optimize feature selection.

Additionally, a heatmap analysis highlighted key body joints influencing classification outcomes. The proposed heatmap is generated through a comparative analysis of feature ranking techniques applied to the biomechanical data from the UI-PRMD and KIMORE benchmark datasets. These datasets provide detailed joint position and motion data, which are more directly aligned with our aim of assessing joint movements and their relevance to rehabilitation exercises.

Ensemble models demonstrated robust performance across multiple datasets, exhibiting low sensitivity to feature ranking methods. In contrast, non-ensemble models showed varying accuracy depending on the feature ranking technique employed, underscoring the impact of feature selection on classification outcomes.

Among the evaluated methods, the Extra Tree Classifier, when paired with the ReliefF feature ranking technique, attained the highest classification accuracy, specifically for the UIPRMD dataset. However, its performance showed variability when applied to the KIMORE dataset. As a result, we recommend adopting FCBF alongside Extra Tree, which demonstrated robust performance across diverse datasets. This approach achieved 99.64% accuracy on the UIPRMD dataset (the second-best result) and 81.85% on the KIMORE dataset (the highest accuracy observed). FCBF consistently delivered stable results across different classifiers, maintaining an average accuracy of 92.65%.

7. Future work

In future work, this study can extend beyond filter-based feature selection techniques to incorporate wrap-based methods, thereby exploring a broader spectrum of feature selection strategies. Future work could also explore the inclusion of additional features and ranking methods, but we believe that the selected techniques provide a comprehensive and balanced analysis for this study. Additionally, the investigation could encompass advanced deep learning architectures such as Convolutional Neural Networks (CNNs), Long Short-Term Memory Networks (LSTMs), and their variants. These models offer the potential for enhanced performance but may pose computational challenges, necessitating the exploration of model fusion techniques to mitigate computational costs and optimize predictive accuracy. Given the constraints of limited dataset sizes, transfer learning emerges as a viable approach to leverage pre-trained models and enhance generalization capabilities. Furthermore, adopting a Synergistic Approach holds promise for integrating diverse feature ranking techniques and classifiers synergistically to capitalize on their collective strengths, thereby advancing the effectiveness and robustness of the proposed methodologies beyond individual components.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Moamen Zaher http://orcid.org/0009-0004-8560-4563
Amr S. Ghoneim http://orcid.org/0000-0003-3522-4875
Laila Abdelhamid http://orcid.org/0000-0002-7928-5680
Ayman Atia http://orcid.org/0000-0003-4998-5624

References

Ahad, M., Antar, A., & Shahid, O. (2019). Vision-based action understanding for assistive healthcare: A short review. In *CVPR Workshops* (pp. 1–11). IEEE.

Almasi, S., Ahmadi, H., Asadi, F., Shahmoradi, L., Arji, G., Alizadeh, M., & Kolivand, H. (2022). Kinect-based rehabilitation systems for stroke patients: A scoping review. *BioMed Research International*, *2022*(1), Article 4339054. https://doi.org/10.1155/bmri.v2022.1

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). IJCAI Organization.

Capecci, M., Ceravolo, M. G., Ferracuti, F., Iarlori, S., Monteriu, 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.7333

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, Y., Huang, S., Yuan, T., Qi, S., Zhu, Y., & Zhu, S.-C. (2019). Holistic++ scene understanding: Single-view 3D holistic scene parsing and human pose estimation with human-object interaction and physical commonsense. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 8648–8657). IEEE.

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

Duarte, N., Rakovic, M., Marques, J., & Santos-Victor, J. (2018). Action alignment from gaze cues in human–human, and human–robot interaction. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Herath, S., Harandi, M., & Porikli, F. (2017). Going deeper into action recognition [1][2]: A survey. *Image and Vision Computing*, *60*, 4–21. https://doi.org/10.1016/j.imavis.2017.01.010

Huang, H.-H., Xu, T., & Yang, J. (2014). Comparing logistic regression, support vector machines, and permanental classification methods in predicting hypertension. In *BMC Proceedings* (Vol. 8, pp. 1–5). Springer.

Im, D., Ma, H., Taylor, G., & Branson, K. (2018). Quantitatively evaluating gans with divergences proposed for training. arXiv:1803.01045.

Lei, Q., Du, J., Zhang, H., Ye, S., & Chen, D. (2019). A survey of vision-based human action evaluation methods. *Sensors*, *19*(19), 4129. https://doi.org/10.3390/s19194129

Liao, Y., Vakanski, A., & Xian, M. (2020). A deep learning framework for assessing physical rehabilitation exercises. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(2), 468–477. https://doi.org/10.1109/TNSRE.2020.2966249

Liao, Y., Vakanski, A., Xian, M., Paul, D., & Baker, R. (2020). A review of computational approaches for evaluation of rehabilitation exercises. *Computers in Biology and Medicine*, *119*, Article 103687. https://doi.org/10.1016/j.compbiomed.2020.103687

Lui, Y. (2012, November 1). Human gesture recognition on product manifolds. *The Journal of Machine Learning Research*, *13*, 3297–321.

Lusa, L. (2017). Gradient boosting for high-dimensional prediction of rare events. *Computational Statistics & Data Analysis*, *113*, 19–37. https://doi.org/10.1016/j.csda.2016.07.016

Machado, M. R., Karray, S., & de Sousa, I. T. (2019). Lightgbm: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. In *2019 14th International Conference on Computer Science & Education (ICCSE)* (pp. 1111–1116). IEEE.

Miron, A., & Grosan, C. (2021). Classifying action correctness in physical rehabilitation exercises. arXiv preprint arXiv:2108.01375.

Nagarajan, T., Feichtenhofer, C., & Grauman, K. (2019). Grounded human- object interaction hotspots from video. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 8688–8697). IEEE.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*, 21. https://doi.org/10.3389/fnbot.2013.00021

Niu, Y., Zhang, J., Zhang, L., & Wang, X. (2023). Deeppose: A deep learning framework for pose-based rehabilitation exercise assessment. *IEEE Transactions on Biomedical Engineering*, *60*(2), 441–451.

Ongun, M. F., Güdükbay, U., & Aksoy, S. (2020). Recognition of occupational therapy exercises and detection of compensation mistakes for cerebral palsy. *Journal of Visual Communication and Image Representation*, *73*, Article 102970. https://doi.org/10.1016/j.jvcir.2020.102970

Palimkar, P., Bajaj, V., Mal, A., Shaw, R., & Ghosh, A. (2022). Unique action identifier by using magnetometer, accelerometer and gyroscope: KNN approach. In *Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021* (pp. 607–631). Springer Singapore.

Patsadu, O., Nukoolkit, C., & Watanapa, B. (2012). Human gesture recognition using kinect camera. In *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)* (pp. 28–32). IEEE.

Pedraza-Hueso, M., Martín-Calzón, S., Díaz-Pernas, F. J., & Martínez-Zarzuela, M. (2015). Rehabilitation using kinect-based games and virtual reality. *Procedia Computer Science*, *75*, 161–168. https://doi.org/10.1016/j.procs.2015.12.233

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing*, *28*(6), 976–990. https://doi.org/10.1016/j.imavis.2009.11.014

Rashid, F. A. N., Suriani, N. S., Mohd, M. N., Tomari, M. R., Zakaria, W. N. W., & Nazari, A. (2020). *Deep convolutional network approach in spike train analysis of physiotherapy movements* (pp. 159–170). Springer.

Réby, K., Dulau, I., Dubrasquet, G., & Aimar, M. (2023, January). Graph transformer for physical rehabilitation evaluation. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)* (pp. 1–8). IEEE.

Sabale, A., & Vaidya, Y. (2016). Accuracy measurement of depth using kinect sensor. In *2016 Conference on Advances in Signal Processing (CASP)* (pp. 155–159). IEEE.

Sardari, F., Paiement, A., Hannuna, S., & Mirmehdi, M. (2020, September). Vi-net–the view-invariant quality of human movement assessment. *Sensors*, *20*(18), 5258. https://doi.org/10.3390/s20185258

Sokolova, M., & Japkowicz, N. (2006). Evaluation metrics for imbalanced classification. In *Australasian Joint Conference on Artificial Intelligence* (pp. 295–304). Springer.

Tamla, P., Hartmann, B., Nguyen, N., Kramer, C., Freund, F., & Hemmje, M. (2022). CIE: A cloud-based information extraction system for named entity recognition in AWS, Azure, and medical domain. In *International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management* (pp. 127–148). Springer.

Tang, Y., Wang, B., He, W., & Qian, F. (2022, February). Pointdet++: An object detection framework based on human local features with transformer encoder. *Neural Computing and Applications* (pp. 1–2). Springer.

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

Urbanowicz, R., Meeker, M., La Cava, W., Olson, R., & Moore, J. (2018). Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*, *85*, 189–203. https://doi.org/10.1016/j.jbi.2018.07.014

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.

Yadav, R. K., Neogi, S. G., & Semwal, V. B. (2023). Human activity identification system for video database using deep learning technique. *SN Computer Science*, *4*(5), 600. https://doi.org/10.1007/s42979-023-02031-5

Yu, L., & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *The Journal of Machine Learning Research*, *5*, 1205–1224.

Zaher, M., Ghoneem, A., & Abdelhamid, L. (2023, April). Comparative study between machine learning algorithms and feature ranking techniques on ui-prmd dataset. PREPRINT (Version 1) available at Research Square

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). IEEE.

Zaher, M., Ghoneim, A. S., Abdelhamid, L., & Atia, A. (2025). Fusing CNNs and attention-mechanisms to improve real-time indoor Human Activity Recognition for classifying home-based physical rehabilitation exercises. Computers in Biology and Medicine, 184, 1093399. https://doi.org/10.1016/j.compbiomed.2024.109399

Zhang, W., Su, C., & He, C. (2020). Rehabilitation exercise recognition and evaluation based on smart sensors with deep learning framework. *IEEE Access*, *8*, 77561–77571. https://doi.org/10.1109/Access.6287639

Zhang, S., Wang, C., Dong, W., & Fan, B. (2022, October). A survey on depth ambiguity of 3D human pose estimation. *Applied Sciences (Basel)*, *12*(20), Article 10591. https://doi.org/10.3390/app122010591

Zhang, Y., Wang, L., Wang, X., & Wang, L. (2020). 3D-resnet: A hierarchical 3D residual network for skeleton-based action recognition. *IEEE Transactions on Cybernetics*, *50*(10), 3819–3831.