



# Deep ensemble learning approach for lower limb movement recognition from multichannel sEMG signals

Pratibha Tokas<sup>1</sup> · Vijay Bhaskar Semwal<sup>1</sup> · Sweta Jain<sup>1</sup>

Received: 27 May 2023 / Accepted: 14 January 2024 / Published online: 17 February 2024  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

## Abstract

Walking is a complex task that requires consistent practice to master, and it involves the synchronisation between the lower limbs and the brain, making it challenging. While bipedal robots have been developed to mimic human walking, they must achieve an efficient gait due to structural differences and walking challenges. This study aims to produce a more human-like walk by analysing human lower extremity activities. To capture the bipedal robot locomotion learning process, an ensemble classifier based on deep learning is introduced to recognise human lower activities. A publicly available UC Irvine Machine Learning Repository (UCI) dataset on surface electromyography (sEMG) signal for the lower extremity of 11 fit participants and 11 participants with knee disorders for sitting while performing knee extension, walking, and standing while performing knee flexion is used. A hybrid ensemble of deep learning models comprising long short-term memory and convolution neural network is employed to classify activities, with reported average accuracies of 98.8%, 98.3%, and 99.3% for healthy subjects for sitting, standing and walking, respectively. Moreover, the ensemble model reported average accuracies of 98.2%, 98.1%, and 99.0% for individuals with knee pathology. Notably, this study holds promising significance, as it has yielded a considerable enhancement in performance as opposed to state-of-the-art work. The applications of this work are diverse and include improving postural stability in elderly subjects, aiding in the rehabilitation of patients recovering from stroke and trauma, generating walking trajectories for robots in complex environments, and reconstructing walking patterns in individuals with impairments.

**Keywords** Human activity recognition (HAR) · sEMG · Deep learning · Hybrid ensemble classifier · Signal processing · Bipedal robots

## 1 Introduction

Gait analysis is crucial in rehabilitation because it gives reliable information regarding a patient's walking pattern and lower limb motions. This information can be used to identify divergences from normal gait and assess rehabilitation interventions' effectiveness. By examining gait, clinicians can demarcate the degree of impairment, monitor

progress during rehabilitation, and adjust the treatment plan accordingly. Gait analysis plays a crucial role in rehabilitation by providing valuable insight into the patient's operational abilities and enabling clinicians to optimise the treatment plan to attain the most suitable outcomes. Injuries like knee osteoarthritis, sciatica, meniscus and anterior cruciate ligament (ACL) are among the leading sources of impairment worldwide, affecting people of all ages [1, 2]. It has been demonstrated that assistive technology has the potential to enhance the standard of life for individuals with these injuries, particularly by tracking their rehabilitation progress [3, 4]. Gait analysis is a standard diagnostic tool for neuromuscular and skeletal disorders that helps classify and assess lower extremity motion [5]. However, conventional clinical rehabilitation techniques involving gait analysis require comprehensive laboratory settings, which can be time-consuming, costly, and inconvenient,

✉ Pratibha Tokas  
213112001@stu.manit.ac.in

Vijay Bhaskar Semwal  
vsemwal@manit.ac.in

Sweta Jain  
Swetajain@manit.ac.in

<sup>1</sup> Maulana Azad National Institute of Technology, Bhopal, M.P., India

especially in secluded areas [6]. Remote monitoring of rehabilitation improvement using wearable devices has become critical to overcoming these limitations. These wearables can not only control assistive devices like exoskeletons but also yield improvement feedback to users and assist clinicians in assessing and treating patients [7–9]. Lower extremity motions are integral to many human activities, including sitting, standing, stair ascent and descent, and squatting. Gait analysis, which involves classifying and evaluating lower leg motions [3, 10], is crucial for diagnosing neuromuscular and skeletal disorders. However, traditional gait analysis techniques require extensive laboratory setups, making it necessary to develop more straightforward methods for assessing gait dysfunction. Various noninvasive and kinematics techniques have been proposed for this purpose [11].

In recent years, electromyography (EMG) has emerged as a widely used approach for recording muscle activities in the skeletal muscles, which is valuable for investigating neuropathic and myopathic conditions, controlling prosthetic devices, and aiding in rehabilitation. EMG signals can be obtained using surface or concentric needle electrodes. sEMG is mainly used in rehabilitation and prosthetic applications. In contrast, concentrically arranged needles diagnose neuromuscular disorders affecting motor units (MUs). In [12], Gautam et al. have used a transfer learning-based LRCN model on the publicly available UCI dataset to predict joint angles and classify lower extremity activities. They have achieved a mean classification accuracy of 92.4% and 98.1% for participants with knee disorders and fit participants, respectively. A general deep-

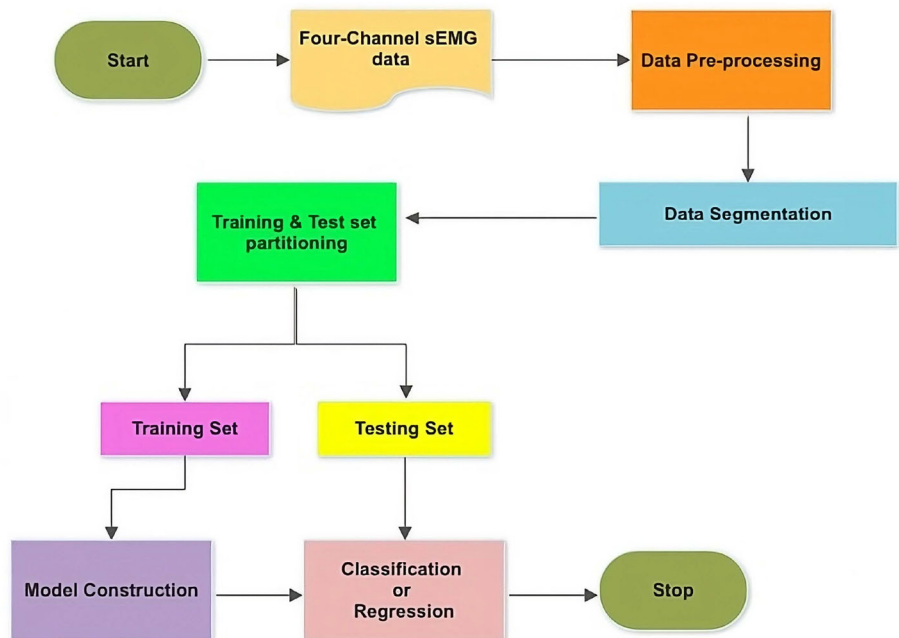
learning approach is shown in Fig. 1. In [13], the authors have generated a dataset using a Microsoft Kinect V2 sensor for 12 different human activities. They have used a hybrid deep learning model for classification, and an average accuracy of 90.89% has been achieved.

Imbalanced class distribution is a significant problem in medical datasets, where the number of samples in different classes varies greatly. This imbalance can lead to prejudiced results towards the majority class, negatively impacting diagnosis accuracy [14]. Therefore, balancing the data by either oversampling the minority class or undersampling the majority class is essential to enhance diagnostic success. In a previous study [15], Rajesh et al. used the AdaBoost ensemble classifier to classify five groups of heartbeats with imbalanced ECG beats. Other studies have also shown that oversampling techniques, such as the Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Technique (ADASYN), can address the class disproportion problem [16–19]. Taft et al., for example, used SMOTE to improve the classification model's performance in identifying adverse medication events in females hospitalised for childbirth and labour. [17].

The major contributions of the authors are ensue:

1. **Data Standardisation:** The data were scaled using a standard scaler, which rescales the features to have a zero mean and unit variance. It ensures that all features have the same scale and distribution, improving the model's efficiency and convergence.

**Fig. 1** Flowchart of general deep learning model



2. **Managing the imbalanced dataset:** The dataset is highly imbalanced. We used ADASYN to balance the dataset.
3. **Data representation:** The four-channel sEMG input signal data were segmented into a 256ms window size with an overlap of 64ms.
4. **Model Design:** Our proposed hybrid ensemble deep neural network architecture for lower extremity recognition utilises an ensemble of CNN and LSTM models trained on four channels of sEMG data. By combining CNN and LSTM, the model can effectively capture human activity data's spatial and temporal aspects.
5. **Performance analysis:** The model has been validated rigorously by investigating the sEMG data and has obtained an average accuracy of 98.8%, 98.3%, and 99.3% for fit individuals for sitting, standing and walking, respectively. Furthermore, an average accuracy of 98.2%, 98.1%, and 99.0% for individuals with knee pathology is a remarkable improvement over the previously published state-of-the-art work.
6. **Statistical Analysis:** The proposed ensemble model has been statistically tested to be significantly different from all competing algorithms using **Friedman test, Bonferroni-Dunn test and Wilcoxon-signed rank test.**

An ensemble deep learning model comprising CNN-LSTM using four-channel sEMG signal data has been proposed. Our hybrid model utilises both CNN and LSTM to capture human activity data's spatial and temporal characteristics obtained through sEMG. While CNN captures spatial information, LSTM captures temporal information in the data.

Moreover, the proposed work dealt with the imbalanced dataset by employing **Adaptive Synthetic Sampling (ADASYN)** technique. This data augmentation technique addresses the problem of imbalanced data by creating synthetic data for classes with fewer samples. ADASYN uses weight distribution while generating synthetic samples for minority classes. Unlike SMOTE, where the synthetic samples are generated uniformly for all minority classes, ADASYN generated synthetic samples according to their difficulty in learning. Hence, more synthetic samples are generated for minority classes that are harder to learn and most likely to be misclassified. The ADASYN algorithm has been shown to improve the performance of classifiers on imbalanced datasets by effectively increasing the size of the minority class and reducing the bias towards the majority class. None of the previous work published on this dataset has dealt with the problem of imbalanced classes, which can make the model overly biased towards the majority class. Moreover, none of the previously published work has performed statistical tests like the **Bonferroni-**

**Dunn test, the Friedman test or Wilcoxon test** to demonstrate whether the proposed hybrid ensemble model is significantly different from the existing models or not. Our proposed model (ensemble) not only outperforms existing competing algorithms but is also proven to be significantly different statistically. The same has been demonstrated in the results (Sect. 6) of this paper.

The remaining portion of this article is laid out as follows. Section 2 comprises the related work of lower extremity activity recognition using wearable sensors. Section 3 contains the description of the individual components of the proposed hybrid ensemble model. Section 4 contains the comprehensive description of the architecture and functionality of our model. Section 5 is the methodology section that entails data collection and pre-processing information. Section 6 outlines the formulas for evaluating the model's performance. It also comprises a detailed description of the results obtained and provides a comparative analysis with other cutting-edge work. Section 6 also gives the results of all statistical tests performed. Lastly, Section 7 comprises the conclusion and the future scope.

## 2 Literature review

Extensive research has been conducted on gait analysis and human activity recognition. The data can be collected as sensor-based or video-based [20]. Both of these methodologies have their benefits and drawbacks. Sensor-based systems are more successful at capturing the tiny subtleties of human motion that may be difficult to see in video frames. Gait analysis can be performed on data obtained from IMU sensors. IMUs typically comprise accelerometers, gyroscopes, and sometimes magnetometers and can be worn on various body parts to capture motion data [10, 21–23]. In [24], Naik et al. investigated 11 fit individuals and 11 with knee disorders. Independent Component Analysis (ICA) separated the sEMG signals into independent components representing individual muscle activations. Six time-domain features were extracted, and an ensemble-based modelling approach was later used for classification. Fit subjects had a mean classification accuracy of 96.1%, while people with knee problems had an accuracy of 86.2%. Zhang et al. introduced a new technique for identifying multi-channel electromyography (EMG) signals using noise-assisted multivariate empirical mode decomposition (NA-MEMD) in [25]. The NA-MEMD technique was used to partition multi-channel EMG data into a set of intrinsic mode functions (IMFs) that capture the signal's numerous frequency components. However, their study was limited to fit participants and was not extended to participants with knee disorders. The

average classification activity obtained was 79% for walking and 83% for both sitting and standing, respectively. In [12], Gautam et al. proposed a combination of convolutional and recurrent neural networks (LRCN) for the recognition of lower extremity movements, prediction of knee joint angles, and classification of lower extremity activities using surface electromyography (sEMG) signals for both fit and participants with known knee disorders. According to [12], the authors reported mean classification accuracy for fit and people with knee ailments of 98.1% and 92.4%, respectively. None of the previous work [12, 24, 25] dealt with a significant problem of an imbalanced dataset. In [26], the authors have proposed a deep neural network model for the consecutive estimation of lower extremity motions using sEMG signals. The model comprises multiple branches, each processing a different segment of the sEMG signal. The outputs from these branches are then combined to produce the final prediction. The model achieves a mean classification accuracy of 90.92% for speed-dependent and 85.4% for speed-independent. In [27], the authors have used the sEMG data of 11 fit participants and 11 participants with known knee disorders. They have denoised the data and obtained time-domain features. Various anomaly detection techniques were used to enhance the model's classification accuracy. Various machine learning classifiers, notably random forest and light gradient boosting, were used to classify the activities. The best classification accuracy of 98.5% was obtained using the iforest anomaly detection algorithm with a light gradient boosting machine. The LDA-PSO-LSTM algorithm combines three techniques: long-short-term memory (LSTM) neural networks, particle swarm optimisation (PSO), and linear discriminant analysis (LDA). To extract discriminant characteristics from sEMG data, the LDA is used. The PSO algorithm is then used to optimise the hyperparameters of the LSTM, which is employed to recognise the gait phases [28]. In [29], authors used a convolutional network for feature selection from the denoised data and a kernel extreme learning machine to classify lower limb activity. The accuracy reported was 95.90% for classification. In [30], the authors have proposed a bimodal hybrid classifier on IMU sensor data for HAR, which can improve the robustness and accuracy of the recognition system. However, the approach may require a relatively large amount of training data to achieve high accuracy. Secondly, the approach may be affected by sensor placement and calibration, which can affect the quality of the motion data.

### 3 Deep neural networks (DNNs)

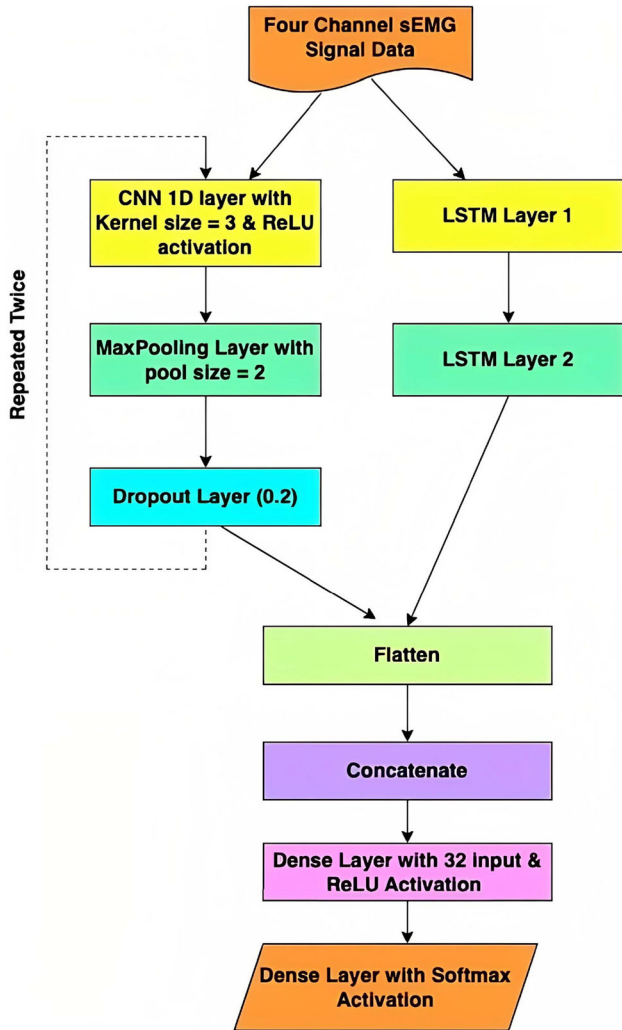
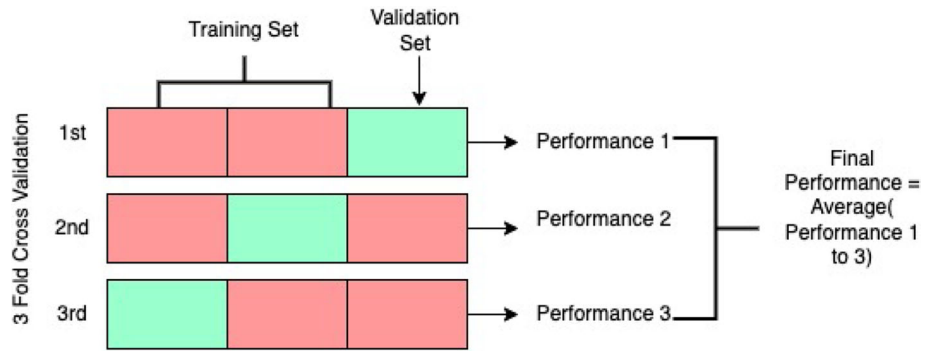
Deep neural network (DNN) is a form of artificial neural network (ANN) that consists of multiple hidden layers between the input and output layers. There are various deep learning models like convolution neural network (CNN), long short-term memory (LSTM) and bi-directional LSTM, etc. In this section, the individual components of the proposed hybrid ensemble deep learning model are detailed.

#### 3.1 Convolution neural network (CNN)

Convolutional neural networks (CNNs) perform various operations such as convolution, pooling, dropout, fully connected layer and activation functions.

1. **Convolution:** The convolution operation involves sliding a filter or kernel over the input data and performing element-wise multiplication, followed by a summation. The equation for the convolution operation is Eq. 1
2. **Pooling:** Pooling is a downsampling operation that reduces the spatial size of the activation maps while preserving the essential features. The two most common types of pooling operations are maximum pooling and average pooling. Maximum pooling selects the maximum value within a window of pixels, while average pooling calculates the average value within the window. In our proposed algorithm, maximum pooling has been employed.
3. **Activation Function:** An activation function adds nonlinearity to the network after each convolution or pooling operation. The most popular activation function is the rectified linear unit, denoted as ReLU, which returns a maximum of 0 or the input value. ReLU and softmax activation functions have the equations Eqs. 2 and 3, respectively.
4. **Fully Connected Layers:** After several convolutional and pooling layers, the feature maps are flattened into a vector and passed through one or more fully connected layers. These layers are similar to feed-forward neural networks, where each neuron is connected to all neurons in the previous layer. The output of the final dense layer is then supplied to a softmax layer for classification. The equation for the same is Eq. 4.
5. **Dropout** Throughout the training process, dropout neurons are selected at random and eliminated by setting their synaptic values to zero in every layer. The pace at which these synaptic are lost is known as the dropout rate. Additionally, it speeds up a model's learning process and inhibits overfitting.

**Fig. 2** Threefold cross-validation architecture



**Fig. 3** Architecture of our proposed model

$$c(t_s) = (a * b)(t_s) \tag{1}$$

$$\text{ReLU}(x) = \max(0, x) \tag{2}$$

$$\text{softmax}(z_i) = \frac{\exp(z_i)}{\sum_{j=1}^K \exp(z_j)} \tag{3}$$

$$X = Wz + \text{bias} \tag{4}$$

where  $c(t_s)$  denotes the convolution of  $a$  and  $b$ , depicted by  $(a * b)(t_s)$ , and is defined as the integral of the product of the two functions, shifted by a parameter  $t_s$ . Furthermore,  $\exp z_i$  and  $\exp z_j$  denote the standard exponential function for input and output vector, respectively, and  $k$  denotes the number of categories.

### 3.2 Long short term memory (LSTM)

It is commonly employed in analysis of time-series and natural language processing.

1. **Input Gate:** The input gate controls what data from the current input are stored in the memory cell. The input gate's equation is Eq. 5.
2. **Forget Gate:** The forget gate determines which information from the previous memory cell should be discarded. Equation 6 gives the equation for the forget gate.
3. **Output Gate:** It identifies which information from the current input and previous memory cell should be outputted. The equation for the output gate is Eq. 7.
4. **Cell State:** It is the long-term memory of the LSTM. It is updated based on the forget, input and output gates. The equations for the same are Eqs. 8 and 9.
5. **Hidden State:** The hidden state is the short-term memory of the LSTM. It is calculated based on the cell state and the output gate. The equation for the hidden state is shown in Eq. 10.

$$h_{ts} = \sigma(W_{\text{input}} * [h_{ts-1}, x_{ts}] + b_{\text{input}}) \tag{5}$$

$$f_{ts} = \sigma(W_{\text{forget}} * [h_{ts-1}, x_{ts}] + b_{\text{forget}}) \tag{6}$$

$$o_{ts} = \sigma(W_{\text{output}} * [h_{ts-1}, x_{ts}] + b_{\text{output}}) \tag{7}$$

$$c_{ts} = \tanh(W_c [\tilde{h}_{ts-1}, x_{ts}] + b_c) \tag{8}$$

$$c_{ts} = f_{ts} * c_{ts-1} + i_{ts} * \tilde{c}_{ts} \tag{9}$$



$$h_{ts} = o_{ts} * \tanh(c_{ts}) \quad (10)$$

where  $f_{ts}$ ,  $o_{ts}$  and  $h_{ts}$  denote the forget, output and input gate, respectively.  $W$  implies the weight for the corresponding gate ( $g$ ),  $h_{ts}$  denotes the previous LSTM block's output at time step  $ts-1$ ,  $x_{ts}$  denotes the input at current time step  $ts$ , and  $b$  implies the bias for the corresponding gate ( $g$ ).  $c_{ts}$  implies cell state at time step  $ts$ , and  $\tilde{c}_{ts}$  denotes the candidate for cell state at time step  $ts$ .

### 3.3 Hybrid ensemble deep learning

A hybrid deep ensemble model is a model that combines both deep learning and ensemble learning. Basic deep-learning models can learn complex patterns in data but may suffer from overfitting and have difficulty generalising to new data. Ensemble learning, on the other hand, combines multiple models to improve performance by leveraging the strengths of each deep learning model and reducing the risk of overfitting. In a hybrid deep ensemble model, multiple deep learning models are trained on the same data but with different architectures, hyperparameters, or random initialisations. The outputs of these models are then combined. This approach can improve the performance of deep learning models by reducing the risk of overfitting and increasing the diversity of the models. It can also provide more robust predictions and better generalisation of new data. This proposed work uses both LSTM and CNN to create a deep hybrid ensemble model that considers the data's spatial and temporal aspects.

## 4 Proposed model

The ensemble model is made up of LSTM and CNN models that run in tandem. Two convolutional layers precede a maximum pooling layer and a dropout layer in the model design. The convolutional layers employ  $3 \times 1$  filters with increasing filters, commencing with 32 and doubling to 64. Using max pooling layers diminishes the

**Table 1** Model architecture

Layer type	Output shape	No. of parameters
Input Layer (CNN)	(None, 256, 4)	0
Conv1D_1	(None, 254, 32)	416
MaxPooling1D_1	(None, 127, 32)	0
Dropout_1	(None, 127, 32)	0
Conv1D_2	(None, 125, 64)	6208
MaxPooling1D_2	(None, 62, 64)	0
Dropout_2	(None, 62, 64)	0
Flatten (CNN)	(None, 3968)	0
Input Layer (LSTM)	(None, 256, 4)	0
LSTM_1	(None, 256, 32)	4736
LSTM_2	(None, 16)	3136
Concatenate	(None, 3984)	0
Dense	(None, 32)	127520
Dense	(None, 3)	99

dimensionality of the feature maps and extracts the most significant characteristics. By randomly turning a percentage of the input units to 0 during training, the dropout layers help circumvent overfitting by manipulating the network to acquire more robust and generalisable features. The LSTM model is split into two layers, each with 32 and 16 units. Finally, the output of the final convolutional layer is transformed into a vector. The outputs of the LSTM and CNN models are concatenated and supplied into a dense layer consisting of 32 units and an activation function of ReLU. The dense layer's output is then transmitted to the dense layer that follows it, which uses a softmax activation function to generate the final estimations for the three categories. To train and access the model on the training set, threefold cross-validation is employed. The 3-fold cross-validation is depicted in Fig. 2. To avoid overfitting, a premature end of the callback is also used. Figure 3 depicts the architecture. Table 1 contains details regarding the model's architecture.

Furthermore, the approach of the suggested study is shown in Algorithm 1.

**Algorithm 1** Proposed Hybrid Ensemble Learning-based Human Activity Recognition System

**Require:** Raw data collected through four channel sEMG signal.

**Ensure:** Classification accuracy, F1\_score, precision and recall values of different lower extremity activities.

- 1: Pre-process the data: standardise and balance it using the ADASYN algorithm and then segment data into a window of size 256ms with 64ms overlap.
- 2: Prepare the dataset by dividing it into training and validation sets.
- 3: Design the ensemble deep learning model using CNN and LSTM.
- 4: Combine the performance of individual models using an ensemble learning classifier.
- 5: Calculate performance measures for various activities, which include precision, recall, accuracy, and F1-score.

**Output:** The class-wise accuracy, classification report and confusion matrix of the different activities.

**Result:** The proposed hybrid ensemble learning-based system achieves high accuracy, precision, recall, and F1-score for recognizing different lower extremity activities based on the sEMG data.

## 5 Experimental setup and details

The content of this section encompasses a description and pre-processing of the dataset used, as well as a schematic diagram illustrating the proposed architecture and the algorithm.

### 5.1 Data acquisition and setup

This study gathered signal data from 22 male volunteers aged 18 and up. Four sEMG channels and one lower limb goniometer measurement channel are included in the data. The participants were split into two groups: eleven fit people and eleven with knee disease, including one with sciatica pain, four with meniscal rupture and six with an anterior cruciate ligament (ACL) injury. Biometrics Ltd. and Datalog MWX8 with the goniometer SG150B were utilised to obtain data on sEMG and knee joints. To ensure interference-free sampling, the SEMG electrodes are separated by 20 mm, and the input impedance is greater than 10 M ohm. The data capture sampling rate was 1000 Hz, and the range for filtering the data instances was between 20 Hz and 460 Hz. The goniometer is positioned on the outer side of the knee. The sEMG electrode channels were positioned on the rectus femoris (rf), semitendinosus (st), vastus medialis (vm), and biceps femoris (bf). The participants' left and right extremities were selected for fit individuals and those with knee disorders. sEMG data and knee joint angles were collected while the subject engaged in three types of physical tasks: standing while making knee flexion motions, walking at ground level, and sitting while performing knee extension actions. These exercises are commonly performed in daily life and rehabilitative activities and do not necessitate the use of additional weights, dumbbells, or fitness equipment. The database

does not include knee joint angle measurements or the sEMG signal during transitional activities such as sit-to-stand or walk-to-sit.

### 5.2 Data pre-processing

The four-channel sEMG signal data was first standardised to have unit variance, and zero mean to ensure the faster convergence of gradient descent and to preclude the overfitting problem [31]. We also employed the ADASYN algorithm to balance the imbalanced dataset. ADASYN creates synthetic samples of the classes with a smaller number of samples. It was done to ensure the model is impartial and unbiased towards the majority class. Later, the four-channel sEMG signal data were segmented into a 256ms window size with 64 ms overlap using a sliding window approach, as per the research published by [24]. We employed k-fold cross-validation, where  $k = 3$ , to perform an equitable comparison with existing cutting-edge work. To create k-folds, we randomly divided the dataset into k equal subgroups and repeated the process k times. One of the k subsets is designated as the testing or validation set in each fold, whereas the remaining k-1 subsets are designated as the learning set. The final result is a calculation of the average classification accuracy over k-folds.

## 6 Results and discussion

### 6.1 Performance evaluation criteria

The efficiency of the suggested hybrid ensemble model is evaluated using accuracy, F1\_score, recall, and precision. The following equations give the formula used for each of the performance evaluation criteria:

$$\text{Prec} = \frac{\text{Tru\_Pos}}{\text{Tru\_Pos} + \text{False\_Pos}} \quad (11)$$

$$\text{Rec} = \frac{\text{Tru\_Pos}}{\text{True\_Pos} + \text{Fal\_Neg}} \quad (12)$$

$$F1\_score = 2 * \frac{\text{Prec} * \text{Rec}}{\text{Prec} + \text{Rec}} \quad (13)$$

$$\text{Acc} = \frac{\text{Tru\_Pos} + \text{Tru\_Neg}}{\text{Tru\_Pos} + \text{Fal\_Pos} + \text{Fal\_Neg} + \text{Tru\_Neg}} * 100 \quad (14)$$

where Acc denotes accuracy, Prec denotes precision, Rec denotes recall, Tru\_Pos denotes true positive, Tru\_Neg denotes true negative, Fal\_Pos denotes false positive and Fal\_Neg denotes false negative.

Apart from the metrics mentioned above, we have also performed statistical tests, namely Friedman test [32], Bonferroni-Dunn test [33] and Wilcoxon test [34], to compare our proposed ensemble model with other competing algorithms.

### 6.1.1 Comparative analysis

The effectiveness of the suggested model was assessed using the aforementioned equations: Eqs. 11–14. The results obtained have been consolidated in Tables 2 and 3. Here, Tables 2 and 3 show the individual subject analysis for healthy and subjects with knee pathology on three different activities, namely walking, standing and sitting obtained by proposed hybrid ensemble model. However, the point is that none of the previously published state-of-the-art work has dealt with the imbalanced dataset problem. Hence, it would be inappropriate to compare the accuracy, as accuracy alone is not the best metric to evaluate the performance of a model, especially when the class distribution is imbalanced. The average classification accuracy, precision, recall and f1\_score obtained by the proposed hybrid ensemble model are 98.8%, 98.8%, 98.7% and 98.7%, respectively, for all the healthy participants and 98.4%, 98.4%, 98.3% and 98.3% for the participants with knee pathology. Moreover, this paper also includes the participant-wise classification accuracy for fit individuals obtained by the proposed hybrid ensemble model in Table 4. Table 4 also compares the suggested model with

**Table 2** Average precision, accuracy, F1\_score and recall obtained of each healthy subject for each activity

Subject	Precision			Recall			F1_score			Accuracy		
	Walking	Standing	Sitting	Walking	Standing	Sitting	Walking	Standing	Sitting	Walking	Standing	Sitting
1	99.6 ± 0.7	100.0 ± 0	100.0 ± 0	100.0 ± 0	99.6 ± 0.7	100.0 ± 0	99.8 ± 0.3	99.8 ± 0.4	100.0 ± 0	99.6 ± 0.7	100.0 ± 0	100.0 ± 0
2	100.0 ± 0	99.2 ± 1.4	98.3 ± 3.0	98.7 ± 2.3	99.6 ± 0.7	99.1 ± 1.5	99.3 ± 1.2	99.4 ± 1.1	98.7 ± 2.3	100.0 ± 0	99.2 ± 1.4	98.3 ± 3.0
3	98.8 ± 2.1	97.2 ± 4.9	96.2 ± 4.9	98.8 ± 2.1	99.2 ± 0.6	92.7 ± 12.7	98.8 ± 2.1	98.2 ± 2.7	94.3 ± 9.0	98.8 ± 2.1	97.2 ± 4.9	96.2 ± 4.9
4	99.1 ± 1.5	96.3 ± 6.4	98.0 ± 3.5	97.9 ± 3.7	99.6 ± 0.7	95.4 ± 8.0	98.5 ± 2.6	97.9 ± 3.7	96.6 ± 5.9	99.1 ± 1.5	96.3 ± 6.4	98.0 ± 3.5
5	99.5 ± 0.8	99.1 ± 1.6	99.6 ± 0.7	99.1 ± 1.6	100.0 ± 0	99.2 ± 1.5	99.3 ± 1.2	99.5 ± 0.8	99.4 ± 1.1	99.5 ± 0.8	99.1 ± 1.6	99.6 ± 0.7
6	96.8 ± 5.5	94.1 ± 9.4	99.5 ± 0.8	95.8 ± 6.3	97.5 ± 0.7	95.4 ± 7.9	96.3 ± 5.9	95.7 ± 7.0	97.3 ± 4.6	96.8 ± 5.5	94.1 ± 9.4	99.5 ± 0.8
7	100.0 ± 0	99.2 ± 1.4	99.6 ± 0.8	100.0 ± 0	99.6 ± 0.7	99.1 ± 1.5	100.0 ± 0	99.4 ± 1.1	99.3 ± 1.2	100.0 ± 0	99.2 ± 1.4	99.6 ± 0.8
8	100.0 ± 0	98.3 ± 0.0	98.8 ± 2.1	100.0 ± 0	99.2 ± 1.5	97.6 ± 0	100.0 ± 0	98.7 ± 0.8	98.2 ± 1.0	100.0 ± 0	98.3 ± 0.0	98.8 ± 2.1
9	100.0 ± 0	99.7 ± 0.5	98.4 ± 2.8	98.0 ± 2.8	100.0 ± 0	100.0 ± 0	99.0 ± 1.4	99.9 ± 0.3	99.2 ± 1.4	100.0 ± 0	99.7 ± 0.5	98.4 ± 2.8
10	100.0 ± 0	99.3 ± 0.6	99.7 ± 0.6	99.8 ± 0.3	99.6 ± 0.8	99.7 ± 0.3	99.9 ± 0.2	99.5 ± 0.7	99.7 ± 0.4	100.0 ± 0	99.3 ± 0.7	99.7 ± 0.6
11	99.0 ± 1.9	99.2 ± 1.3	98.3 ± 3.0	99.3 ± 1.3	98.1 ± 3.3	99.4 ± 1.1	99.1 ± 1.6	98.7 ± 2.3	98.8 ± 2.1	99.0 ± 1.9	99.2 ± 1.4	98.3 ± 3.0
Average	99.3 ± 1.1	98.3 ± 2.5	98.8 ± 2.0	98.8 ± 1.9	99.3 ± 1.2	98.0 ± 3.1	99.1 ± 1.5	98.8 ± 1.9	98.3 ± 2.6	99.3 ± 1.1	98.3 ± 2.5	98.8 ± 2.0



**Table 3** Average precision, accuracy, F1\_score and recall obtained of each subject with knee pathology for each activity

Subject	Precision			Recall			F1_score			Accuracy		
	Walking	Standing	Sitting	Walking	Standing	Sitting	Walking	Standing	Sitting	Walking	Standing	Sitting
1	99.7 ± 0.5	99.3 ± 1.3	99.4 ± 1.0	99.3 ± 0.2	100.0 ± 0	99.6 ± 0.8	99.5 ± 0.8	99.6 ± 0.6	99.5 ± 0.9	99.7 ± 0.5	99.3 ± 1.3	99.4 ± 1.0
2	99.5 ± 0.8	98.1 ± 3.3	98.0 ± 2.3	100.0 ± 0	97.7 ± 2.9	98.0 ± 3.4	99.8 ± 0.4	98.0 ± 3.1	98.0 ± 2.8	99.5 ± 0.8	98.1 ± 3.3	98.0 ± 2.3
3	99.7 ± 0.4	98.7 ± 1.8	98.8 ± 2.1	99.3 ± 1.3	99.3 ± 0.8	98.4 ± 1.5	99.4 ± 0.5	99.0 ± 1.3	98.6 ± 1.8	99.7 ± 0.4	98.7 ± 1.8	98.8 ± 2.1
4	99.9 ± 0.2	99.9 ± 0.2	98.9 ± 1.1	99.4 ± 0.5	99.9 ± 0.2	99.6 ± 0.7	99.7 ± 0.3	99.9 ± 0.2	99.2 ± 0.9	99.9 ± 0.2	99.9 ± 0.2	98.9 ± 1.1
5	99.6 ± 0.7	98.0 ± 1.5	99.3 ± 0.6	98.6 ± 0.9	99.7 ± 0.4	99.3 ± 1.2	99.1 ± 0.7	98.9 ± 1.0	99.3 ± 0.8	99.6 ± 0.7	98.0 ± 1.5	99.3 ± 0.6
6	98.6 ± 2.5	94.9 ± 8.9	97.9 ± 3.7	94.4 ± 9.7	99.7 ± 0.5	95.5 ± 7.7	96.3 ± 6.4	97.1 ± 5.0	96.7 ± 5.8	98.6 ± 2.5	94.9 ± 8.9	97.9 ± 3.7
7	100.0 ± 0	98.7 ± 2.3	98.4 ± 1.9	100.0 ± 0	98.2 ± 2.1	98.8 ± 2.1	100.0 ± 0	98.4 ± 2.2	98.6 ± 2.0	100.0 ± 0.0	98.7 ± 2.3	98.4 ± 1.9
8	98.4 ± 1.9	96.7 ± 2.9	97.0 ± 5.1	95.0 ± 5.2	98.7 ± 1.3	98.5 ± 2.7	96.6 ± 3.4	97.7 ± 1.6	97.7 ± 3.9	98.4 ± 1.9	96.7 ± 2.9	97.0 ± 5.1
9	96.0 ± 5.5	97.5 ± 3.9	97.4 ± 4.4	98.0 ± 3.0	100.0 ± 0	90.6 ± 14.1	97.0 ± 4.2	98.7 ± 2.0	93.7 ± 9.8	96.0 ± 5.5	97.5 ± 3.9	97.4 ± 4.4
10	99.1 ± 1.6	99.4 ± 1.1	97.0 ± 5.2	96.6 ± 5.8	99.4 ± 1.1	99.4 ± 1.0	97.8 ± 3.8	99.4 ± 1.1	98.2 ± 3.2	99.1 ± 1.6	99.4 ± 1.1	97.0 ± 5.2
11	99.0 ± 1.7	98.3 ± 1.9	97.6 ± 4.1	97.9 ± 3.6	99.7 ± 0.4	97.4 ± 3.4	98.5 ± 2.7	99.0 ± 1.1	97.5 ± 3.7	99.0 ± 1.7	98.3 ± 1.9	97.6 ± 4.1
Average	99.0 ± 1.4	98.1 ± 2.6	98.2 ± 2.9	98.0 ± 2.8	99.3 ± 0.9	97.7 ± 3.5	98.5 ± 2.1	98.7 ± 1.7	97.9 ± 3.2	99.0 ± 1.4	98.1 ± 2.6	98.2 ± 2.9

other cutting-edge work, namely MyoNet [12], ICA-EBM [24], NA-MEMD [25], on the same dataset for fit participants. The average classification accuracy of walking activity obtained by the proposed model is  $99.3\% \pm 1.1$  as opposed to  $98.2\% \pm 1.6$  in [12],  $96.0\% \pm 1.3$  in [24], and  $79.0\%$  in [25]. Furthermore, the average classification accuracy for standing activity for fit participants obtained by the proposed hybrid ensemble model is  $98.3\% \pm 2.5$  and for sitting is  $98.8\% \pm 2.0$  which shows a significant improvement over the previously published state-of-the-artwork:  $97.7\% \pm 1.3$ ,  $98.4\% \pm 1.4$  [12] and  $96.2\% \pm 1.2$ ,  $96.2\% \pm 1.1$  in [24],  $83.0\%$  and  $83.0\%$  in [25], respectively. The comparison of mean classification accuracy obtained using the proposed hybrid ensemble work with other state-of-the-art work for healthy participants presented in Table 4 is further illustrated in Fig. 4a.

Our proposed model also shows promising results for classifying various lower limb activities performed by individuals with knee pathology. Table 3 comprises the class-wise average accuracy, f1\_score, precision, and recall for each individual with known knee pathology. Moreover, Table 5 compares the proposed model's performance on data obtained from participants with knee pathology with

existing cutting-edge work, namely MyoNet [12] and ICA-EBM [24]. The mean accuracy of  $98.2\% \pm 2.9$ ,  $98.1\% \pm 2.6$ ,  $99.0\% \pm 1.4$  for sitting, standing, and walking activities for participants having knee disorders using the proposed hybrid ensemble model. It shows a significant improvement from the average accuracy obtained by previous cutting-edge work;  $86.4\%$ ,  $85.5\%$ ,  $86.6\%$  in [24],  $92.2\%$ ,  $92.3\%$ ,  $92.8\%$  in [12] for walking, standing, and sitting, respectively. The details presented in Table 5 can also be seen in Fig. 4b. The previously published state-of-the-art work has struggled to perform activity classification for subjects with knee pathology, as for such individuals, each pathology is different and can lead to different functional limitations and activity restrictions. Therefore, developing a single classification model that accounts for different pathologies in different individuals is challenging. Our proposed model achieved an average accuracy of  $98.4\%$  across all individuals with pathology as opposed to  $92.4\%$  in [12],  $86.2\%$  in [24]. The proposed model outperforms the previously published state-of-the-art work [24] by  $12.2\%$  and [12] by  $6.0\%$ . Moreover, it is worth noting that in the real world, signals contain noise, and our proposed method works on noisy signals and does not

**Table 4** Comparative analysis of classification accuracy (%) for different methods on the three activity classes for the healthy subjects dataset

Subject	MyoNet [12]			ICA-EBM [24]			NA-MEMD [25]			Our method		
	Walking	Standing	Sitting	Walking	Standing	Sitting	Walking	Standing	Sitting	Walking	Standing	Sitting
1	98.2 ± 1.8	98.5 ± 1.2	97.3 ± 1.8	95.6 ± 1.2	96.3 ± 1.3	96.2 ± 1.3	78.0	76.0	84.0	99.6 ± 0.7	100.0 ± 0	100.0 ± 0
2	97.6 ± 2.1	97.2 ± 1.6	98.6 ± 1.7	96.3 ± 1.3	96.4 ± 1.3	96.4 ± 1.2	79.0	76.0	84.0	100.0 ± 0	99.2 ± 1.4	98.3 ± 3.0
3	97.3 ± 1.7	93.9 ± 1.7	99.2 ± 0.6	95.8 ± 1.4	96.2 ± 1.2	96.1 ± 1.2	83.0	80.0	88.0	98.8 ± 2.1	97.2 ± 4.9	96.2 ± 4.9
4	98.4 ± 0.8	97.3 ± 1.4	99.1 ± 0.2	96.4 ± 1.3	96.3 ± 1.3	96.3 ± 1.1	76.0	76.0	80.0	99.1 ± 1.5	96.3 ± 6.4	98.0 ± 3.5
5	99.1 ± 1.5	99.6 ± 2.2	98.2 ± 2.1	96.1 ± 1.4	95.6 ± 1.1	96.2 ± 1.0	81.0	82.0	76.0	99.5 ± 0.8	99.1 ± 1.6	99.6 ± 0.7
6	98.1 ± 1.6	97.0 ± 1.7	99.5 ± 0.6	95.6 ± 1.6	96.1 ± 1.4	96.3 ± 1.4	84.0	80.0	75.0	96.8 ± 5.5	94.1 ± 9.4	99.5 ± 0.8
7	97.5 ± 2.2	98.5 ± 1.1	98.2 ± 2.1	96.2 ± 1.1	95.8 ± 1.2	96.1 ± 1.1	79.0	89.0	77.0	100.0 ± 0	99.2 ± 1.4	99.6 ± 0.8
8	99.4 ± 0.6	95.7 ± 0.7	98.9 ± 1.6	96.3 ± 1.5	96.2 ± 1.3	96.2 ± 1.4	79.0	89.0	86.0	100.0 ± 0	98.3 ± 0.0	98.9 ± 2.1
9	96.5 ± 2.6	98.6 ± 1.5	97.3 ± 2.6	95.8 ± 1.3	96.4 ± 1.3	96.1 ± 1.2	79.0	92.0	87.0	100.0 ± 0	99.7 ± 0.5	98.4 ± 2.8
10	100 ± 0	99.3 ± 0.2	98.5 ± 1.6	95.9 ± 1.4	96.5 ± 1.2	96.3 ± 1.2	72.0	89.0	91.0	100.0 ± 0	99.3 ± 0.7	99.7 ± 0.6
11	99.0 ± 1.5	99.3 ± 1.3	98.3 ± 1.3	95.8 ± 1.2	96.2 ± 1.5	96.3 ± 1.3	79.0	88.0	86.0	99.0 ± 1.9	99.3 ± 1.4	98.3 ± 3.0
Average	98.2 ± 1.6	97.7 ± 1.3	98.4 ± 1.4	96.0 ± 1.3	96.2 ± 1.2	96.2 ± 1.1	79.0	83.0	83.0	99.3 ± 1.1	98.3 ± 2.5	98.8 ± 2.0

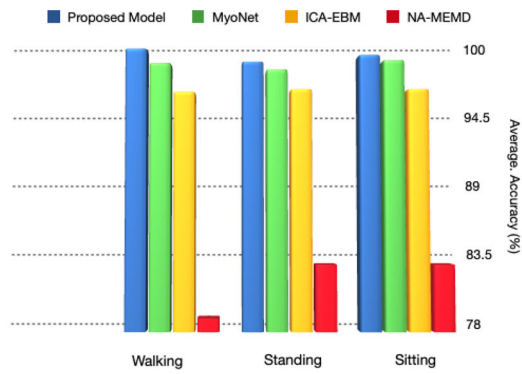
employ any denoising technique. Furthermore, the proposed model showed an average precision, recall and  $f1\_score$  of 98.4%, 98.3% and 98.4%, as opposed to 93.4%, 92.6% and 92.9%, respectively, across subjects with knee pathology. The comparison of the overall performance of the proposed model with the MyoNet method [12] for healthy subjects and subjects with knee pathology has been shown in Fig. 4c and d, respectively. It can be noted that the proposed model outperforms the MyoNet method [12] in terms of recall,  $F1\_score$  and accuracy for healthy subjects and in terms of accuracy, recall, precision and  $F1\_score$  for subjects the knee pathology.

In this research, a subject-wise comparison of the suggested ensemble model with CNN-only and LSTM-only models for healthy subjects and subjects with knee disorder is also shown in Tables 6 and 7, respectively. The mean accuracy for walking, standing, and sitting using the CNN-only model is 98.3%, 96.7% and 96.0% for fit individuals and 97.6%, 97.4% and 97.2%, respectively, for subjects with knee disorders. Similarly, the LSTM-only models achieved an average accuracy of 86.0%, 81.7% and 91.2% for walking, standing, and sitting activities for fit participants and 80.1%, 79.3% and 78.3%, respectively, for

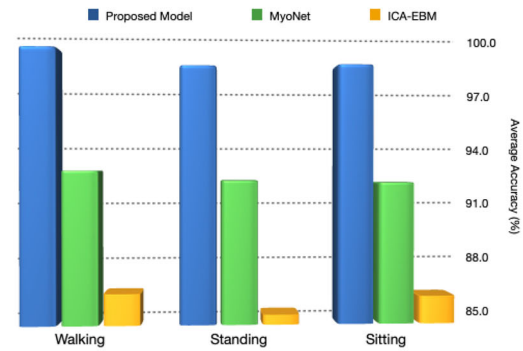
participants with knee disorder. It can be noted that LSTM\_only suffers massively in classifying the lower limb-related activities for both fit and individuals with knee pathology, and CNN\_only performs better than LSTM\_only but the hybrid deep ensemble model performs the best as it captures both the spatial and temporal aspect of the data. The comparison of accuracy obtained using CNN\_only, LSTM\_only and proposed hybrid model have also been further illustrated in Fig. 4e and f.

### 6.1.2 Statistical test

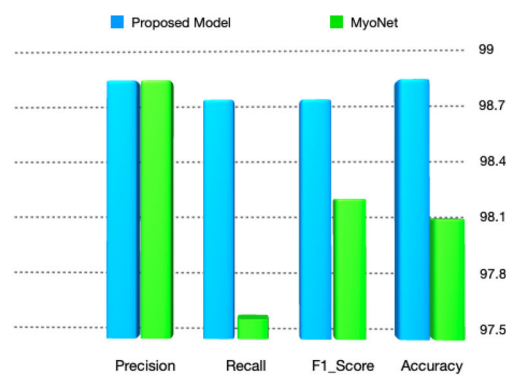
To compare the efficiency of the suggested work with that of other cutting-edge approaches, we have performed **friedman** [32], **bonferroni-dunn** [33] and **Wilcoxon signed rank test** [34] on the proposed ensemble model, ICA [24], MyoNet [12]. In addition to the proposed ensemble model, we have also compared the CNN-only and LSTM-only models. We have not statistically compared our proposed model with NA-MEMD [25] because in [25], the study was limited to healthy subjects, and pathological subjects were not considered.



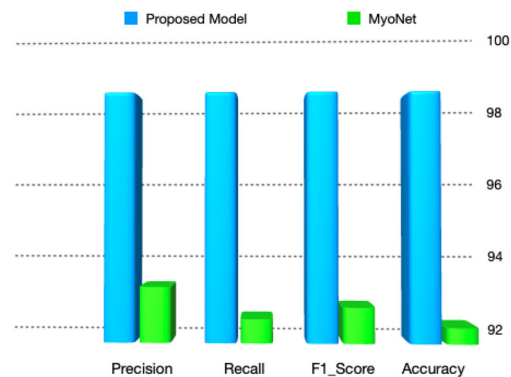
(a) Comparative analysis of the suggested model with NA-MEMD, MyoNet and ICA-EBM for mean classification accuracy of sitting, standing and walking across all healthy participants



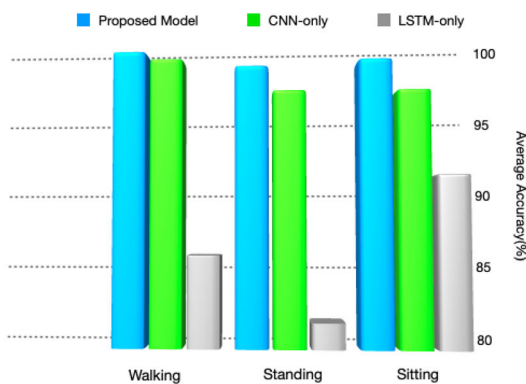
(b) Comparative analysis of the proposed model with MyoNet and ICA-EBM for mean classification accuracy of standing, sitting and walking across all subjects with knee pathology



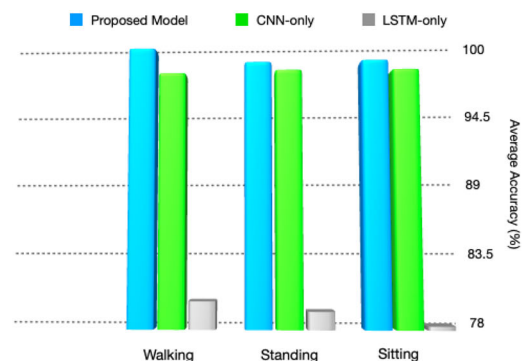
(c) Comparison of overall performance of our proposed model for healthy subjects with MyoNet



(d) Comparative analysis of overall performance of our proposed model for individuals with knee disorder with MyoNet



(e) Comparative analysis of the suggested model with CNN and LSTM for mean classification accuracy of standing, sitting and walking across all healthy subjects



(f) Comparative analysis of the suggested model with CNN and LSTM for mean classification accuracy of standing, sitting and walking across all subjects with knee disorder

**Fig. 4** Comparative analysis of proposed Model with other cutting-edge methodology

The Friedman test is a widely accepted statistical test to compare the performance of multiple models. Let  $r_j^i$  signify the rank of the  $i$ th algorithm over the  $j$ th dataset,  $r_i$  denote

the mean rank of the  $i$ th method, and  $n$  denotes the count of sets or samples.  $k$  signifies the count of competing

**Table 5** Comparative analysis of classification accuracy (%) for subjects with knee pathology

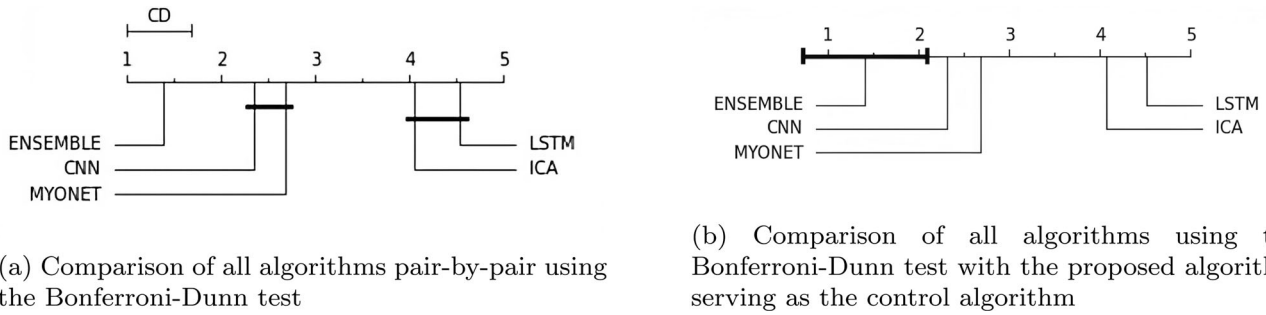
Subject	MyoNet [12]			ICA-EBM [24]			Our method		
	Walking	Standing	Sitting	Walking	Standing	Sitting	Walking	Standing	Sitting
1	92.7 ± 1.8	92.1 ± 1.5	90.7 ± 2.1	87.7 ± 1.3	85.8 ± 1.3	86.4 ± 1.4	99.7 ± 0.5	99.3 ± 1.3	99.4 ± 1.0
2	93.6 ± 2.3	91.6 ± 1.6	92.9 ± 1.2	86.8 ± 1.2	85.2 ± 1.2	86.5 ± 1.2	99.5 ± 0.8	98.1 ± 3.3	98.0 ± 2.3
3	94.2 ± 0.7	93.4 ± 0.4	92.6 ± 1.7	86.4 ± 1.3	85.4 ± 1.3	86.7 ± 1.3	99.7 ± 0.4	98.7 ± 1.8	98.8 ± 2.1
4	92.7 ± 2.1	92.9 ± 1.1	91.8 ± 2.2	86.4 ± 1.4	85.5 ± 1.2	86.4 ± 1.2	99.9 ± 0.2	99.9 ± 0.2	98.9 ± 1.1
5	92.6 ± 2.3	91.3 ± 0.7	92.5 ± 1.8	86.8 ± 1.2	85.4 ± 1.4	86.5 ± 1.6	99.6 ± 0.7	98.0 ± 1.5	99.3 ± 0.6
6	92.9 ± 2.4	91.9 ± 0.9	91.4 ± 0.9	86.3 ± 1.3	85.5 ± 1.1	86.4 ± 1.4	98.6 ± 2.5	94.9 ± 8.9	97.9 ± 3.7
7	93.3 ± 1.8	93.6 ± 2.3	92.9 ± 0.7	86.4 ± 1.4	85.7 ± 1.5	86.3 ± 1.3	100.0 ± 0.0	98.7 ± 2.3	98.4 ± 1.9
8	89.6 ± 2.1	92.5 ± 1.2	92.2 ± 1.81	86.3 ± 1.3	85.4 ± 1.2	86.8 ± 1.2	98.4 ± 1.9	96.7 ± 2.9	97.0 ± 5.1
9	94.1 ± 1.0	87.9 ± 1.7	93.1 ± 0.7	86.7 ± 1.2	85.5 ± 1.4	86.3 ± 1.3	96.0 ± 5.5	97.5 ± 3.9	97.4 ± 4.4
10	93.4 ± 0.6	94.8 ± 1.1	92.6 ± 2.2	86.8 ± 1.5	85.8 ± 1.3	86.2 ± 1.4	99.1 ± 1.6	99.4 ± 1.1	97.0 ± 5.2
11	91.9 ± 2.1	93.3 ± 1.3	92.4 ± 2.4	86.4 ± 1.2	85.5 ± 1.4	86.3 ± 1.2	99.0 ± 1.7	98.3 ± 1.9	97.6 ± 4.1
Average	92.8 ± 1.7	92.3 ± 1.2	92.2 ± 1.6	86.6 ± 1.3	85.5 ± 1.3	86.4 ± 1.3	99.0 ± 1.4	98.1 ± 2.6	98.2 ± 2.9

**Table 6** Comparative analysis of suggested Ensemble model with LSTM-only and CNN-only model for healthy subjects

Subject	Ensemble model			CNN-only			LSTM-only		
	Walking	Standing	Sitting	Walking	Standing	Sitting	Walking	Standing	Sitting
1	99.6	100.0	100.0	99.6	99.6	97.8	98.7	97.8	96.2
2	100.0	99.2	98.3	100.0	98.4	98.7	81.8	93.4	90.2
3	98.8	97.2	96.2	97.9	93.5	92.8	94.2	88.1	93.5
4	99.1	96.3	98.0	99.1	92.9	88.9	89.5	73.6	88.2
5	99.5	99.1	99.6	99.1	99.1	97.9	99.5	86.6	83.5
6	96.8	94.1	99.5	97.8	90.5	93.9	41.8	51.1	86.7
7	100.0	99.2	99.6	96.0	96.2	95.6	76.3	66.05	96.8
8	100.0	98.3	98.9	99.2	98.4	96.8	95.6	77.9	90.36
9	100.0	99.7	98.4	95.8	100.0	98.8	86.3	86.8	93.5
10	100.0	99.3	99.7	97.8	98.9	97.8	96.5	95.9	93.3
11	99.0	99.3	98.3	98.8	96.6	97.5	85.8	81.9	90.7
Average	99.3	98.3	98.8	98.3	96.7	96.0	86.0	81.7	91.2

**Table 7** Comparative analysis of proposed Ensemble model with LSTM-only and CNN-only model for subjects with knee pathology

Subject	Ensemble model			CNN-only			LSTM-only		
	Walking	Standing	Sitting	Walking	Standing	Sitting	Walking	Standing	Sitting
1	99.7	99.3	99.4	98.0	99.3	99.8	96.3	89.3	91.7
2	99.5	98.1	98.0	99.5	96.4	95.7	68.9	96.9	32.5
3	99.7	98.7	98.8	99.7	97.5	99.4	96.8	86.5	80.5
4	99.9	99.9	98.9	94.1	99.8	98.2	85.3	76.72	74.7
5	99.6	98.0	99.3	99.2	95.3	99.3	94.9	90.7	79.6
6	98.6	94.9	97.9	97.1	97.6	97.6	61.0	83.9	66.9
7	100.0	98.7	98.4	99.2	97.4	97.5	82.5	66.5	93.6
8	98.4	96.7	97.0	98.7	95.1	95.7	88.1	62.4	80.1
9	96.0	97.5	97.4	92.8	97.3	93.2	67.1	65.6	80.2
10	99.1	99.4	97.0	97.6	98.4	95.9	73.3	84.9	82.5
11	99.0	98.3	97.6	97.9	97.8	97.2	66.4	69.5	99.4
Average	99.0	98.1	98.2	97.6	97.4	97.2	80.1	79.3	78.3



**Fig. 5** Comparative analysis of each algorithm using the Bonferroni-Dunn test

methods. In the event of a tie between any algorithms, the average rank is shared.

The Friedman statistic ( $F_r$ ) is evaluated using the Eqs. [15] and [16].

$$F_r = \frac{(n - 1)\chi_f^2}{n(k - 1) - \chi_f^2} \tag{15}$$

where

$$\chi_f^2 = \frac{12n}{k(k + 1)} \left( \sum_{i=1}^k r_i^2 - \frac{k(k + 1)^2}{4} \right) \tag{16}$$

In our case,  $n = 66$ ,  $k = 5$ , chosen level of significance ( $\alpha$ ) = 0.05.

$$\chi_f^2 = \frac{12 \times 66}{5 \times 6} \times \left( 51.529 - \frac{5 \times (6)^2}{4} \right) \tag{17}$$

$$\chi_f^2 = 172.367 \tag{18}$$

$$F_r = \left( \frac{65 \times 172.367}{66 \times 4 - 172.367} \right) = 122.268 \tag{19}$$

Critical value  $p$ -value =  $1.1415 \times 10^{-36}$

This  $p$ -value indicates decisive statistical significance, as it is much smaller than the threshold for statistical significance ( $\alpha$ ) = 0.05. This means we can reject the null hypothesis that the models are not statistically different with high confidence. Additionally, we have used the pairwise comparison method of the Bonferroni-Dunn test [33] to determine whether algorithms are significantly different. Using the proposed model as the governing

algorithm, we have also examined whether it performs noticeably better than rival techniques. Equation 20 determines the critical distance (CD).

$$CD = x_\alpha \left[ \frac{k(k + 1)}{6n} \right]^{1/2} \tag{20}$$

At the level of significance,  $\alpha = 0.05$ , the obtained CD = 0.6876. The result for the Bonferroni-Dunn test with the proposed algorithm as a control algorithm and a general comparison is shown in Fig. 5a and b, respectively. Figure 5a shows that the CNN-only model and MyoNet [12] are similar and not statistically different, and LSTM-only and ICA [24] are also similar to each other and not statistically significantly different. As depicted in Fig. 5a and b, it is further emphasised that the proposed ensemble model (referred to as ensemble) is not only better performing than all the other competing algorithms in terms of accuracy but is also significantly different from them.

To further validate that the proposed hybrid ensemble model outperforms the other model and is significantly different from existing state-of-the-art work, we have employed **Wilcoxon Signed-Rank Test** [34]. The result of the Wilcoxon signed-rank test has been shown in Table 8. As shown in Table 8, the  $p$  – value is significantly less than the significance level  $\alpha = 0.05$ . The proposed ensemble model is significantly different from other competing algorithms and outperforms each of them.

**Table 8** Wilcoxon signed-rank test results for the suggested ensemble model opposed to other models

S. No.	Model	Test statistic	$P$ value	Result
1	Ensemble vs MyoNet [12]	109.5	$7.67 \times 10^{-10}$	Different
2	Ensemble vs ICA [24]	6.5	$3.24 \times 10^{-12}$	Different
3	Ensemble vs CNN-only	178.5	$9.63 \times 10^{-08}$	Different
4	Ensemble vs LSTM-only	3.0	$2.77 \times 10^{-12}$	Different



**Table 9** Comparison with related work

Parameters	[38]	[24]	[25]	[39]	[12]	Our model
Dealt with class imbalance	No	No	No	No	No	Yes
Human intervention in feature selection	Yes	Yes	Yes	Yes	No	No
Number of different features extracted	39	7	16	6	Data-driven feature extraction	Data-driven feature extraction
Walking accuracy (%)	88	91.3	79.0	–	95.5	99.2
Sitting accuracy (%)	94	91.3	83.0	–	95	98.2
Standing accuracy (%)	92	90.8	83.0	–	95.3	98.5
Average accuracy (%)	91.3	91.1	81.6	85	95.2	98.6
Statistical comparison	No	No	No	No	No	Yes

## 6.2 Discussion

Individuals with knee-related illnesses like osteoarthritis, meniscus tears, and ACL tears frequently struggle with daily activities involving lower extremity motions such as walking, standing, and sitting [35]. Surface electromyography signals obtained from the muscles, like the quads and hamstrings, during motion can aid in diagnosis by clinicians and assist in rehabilitation [36, 37]. Additionally, sEMG signals can help evaluate the improvement in physiotherapy sessions when using a network-based rehabilitation approach.

We have suggested a hybrid ensemble model comprising LSTM and CNN that captures the spatial and temporal aspects of the signal. In [25], Zhang et al. investigated the use of intrinsic mode functions (IMFs) obtained from the decomposition of surface electromyography (sEMG) signals as features for identifying the lower leg motion. They explored the use of three different decomposition methods, namely multivariate empirical mode decomposition (MEMD), noise-assisted MEMD (NA-MEMD) and empirical mode decomposition (EMD), to extract the IMFs from the raw sEMG signal. One of the significant limitations of their work is that they have restricted the scope of their study to the data of healthy subjects and not extended it to subjects with knee disorders. In [24], Naik et al. have extracted six time-domain features and then applied feature dimensionality reduction. Furthermore, the authors used SVM with an RBF kernel to classify lower limb motion in fit subjects and subjects with knee abnormalities. We have compared our suggested ensemble model with previous cutting-edge work and CNN-only and LSTM-only models. The ensemble model outperforms all the other methods with an average accuracy of 99.3%, 98.3% and 98.8% for a walk, stand and sit activity for healthy subjects and 99.0%, 98.1% and 98.2% respectively, for subjects with knee pathology. Moreover, our proposed ensemble model is also statistically proven to differ significantly from all the competing methodologies. For better understanding, the

comparative analysis of the previous cutting-edge techniques with the proposed ensemble model is shown in Table 9.

## 7 Conclusion and future scope

A hybrid ensemble deep learning model is proposed for classifying lower extremity movement. The proposed hybrid ensemble model significantly improves the recall, f1\_score, accuracy, and precision values over the previously published state-of-the-art work. Our proposed hybrid ensemble model provides an average accuracy of 98.8%, precision of 98.8%, f1\_score of 98.7% and recall of 98.7%. In addition to the proposed hybrid ensemble model comprising CNN and LSTM models, this work has also standardised the data zero mean and unit variance that leads to faster convergence of the models along with enhancing model's efficiency. The dataset was highly imbalanced, and the previous works had not considered that aspect. An imbalanced dataset may lead to discriminatory behaviour by the model towards the majority class, and hence accuracy may not be the best metric to gauge the model's performance. We have handled the imbalanced dataset problem using the adaptive synthetic oversampling algorithm (ADASYN). Enhancing classification accuracy, our suggested hybrid ensemble model beats all current cutting-edge work, particularly for participants with knee pathologies. For fit participants and participants with knee disorders, our model had a mean classification accuracy of 98.8% and 98.4%, respectively. Additionally, the proposed ensemble model has been also compared with CNN and LSTM individually. And our results show that the proposed hybrid ensemble models works better in catering both temporal and spacial aspect of the data. Moreover, the proposed work has also incorporated statistical tests like the Friedman, Bonferroni-Dunn, and Wilcoxon tests to compare our proposed model with all other competing algorithms. All the statistical tests and performance metrics

proved that our model is superior and significantly different from all other competing algorithms discussed. None of the previous works has standardised the data, handled the imbalanced dataset problem, or compared their algorithms statistically with previous state-of-the-art work. All the points mentioned above have been addressed in the proposed work. In the future, we intend to create a model that can classify activities and predict joint angles. The proposed model can also be validated using a larger dataset. We will additionally employ TinyML to significantly reduce the model's complexity and make it more computationally efficient.

**Acknowledgements** This work was supported by the Ministry of Education, Government of India, through the project HEFA CSR under Grant SAN/CSR/08/2021-22. The authors would like to thank the volunteers who have participated and contributed in preparation of the dataset and also the creator of the dataset.

**Author contributions** P.T and V.B contributed to conceptualisation and methodology; P.T contributed to software; P.T, V.B and S.J contributed to validation and writing—original draft preparation and review and editing; V.B and S.J supervised the study. All authors have read and approved the final manuscript.

**Funding** The work is funded by Ministry of Education, Govt. of India to Dr. Vijay Bhaskar Semwal under Higher Education Financing Agency (HEFA) under CSR Grant with Sanctioned order no- SAN/CSR/08/2021-22.

**Data availability** The dataset analysed during the current study are available in the UCI Machine Learning repository, <http://archive.ics.uci.edu/ml/datasets/emg+dataset+in+lower+limb>.

## Declarations

**Conflict of interest** The authors state that they do not have any conflicts of interest.

**Compliance with Ethical Standards** All the ethical issues have been taken care of while writing the manuscript, and we have complied with all the standards to the best of our knowledge.

## References

- Kujala UM, Orava S, Parkkari J, Kaprio J, Sarna S (2003) Sports career-related musculoskeletal injuries: long-term health effects on former athletes. *Sports Med* 33:869–75
- Kianifar R, Lee A, Raina S, Kulic D (2017) Automated assessment of dynamic knee valgus and risk of knee injury during the single leg squat. *IEEE J Trans Eng Health Med* 30(5):1–3
- Dua Nidhi, Singh Shiva, Semwal Vijay, Challa Sravan (2022) Inception inspired CNN-GRU hybrid network for human activity recognition. *Multimed Tools Appl* 82:03
- Anjali Gupta and Vijay Bhaskar Semwal (2022) Occluded gait reconstruction in multi person gait environment using different numerical methods. *Multimed Tools Appl* 81(16):23421–23448
- Baker R (2006) Gait analysis methods in rehabilitation. *J Neuroeng Rehabil* 3(1):4
- Muro-De-La-Herran A, Garcia-Zapirain B, Mendez-Zorrilla A (2014) Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors Basel Switz* 14:3362–94
- Semwal VB, Gaud N, Nandi GC. (2019) Human gait state prediction using cellular automata and classification using elm. In M. Tanveer and Ram Bilas Pachori, editors, *Machine intelligence and signal analysis*, pp 135–145, Singapore, Springer Singapore
- Semwal VB, Gupta A, Lalwani P (2021) An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition. *J Supercomput* 77(11):12256–12279
- Challa SK, Kumar A, Semwal VB (2021) A multibranch CNN-BILSTM model for human activity recognition using wearable sensor data. *Vis Comput* 38(12):4095–4109
- Qiu Sen, Fan Tianqi, Jiang Junhan, Wang Zhelong, Wang Yongzhen, Junnan Xu, Sun Tao, Jiang Nan (2023) A novel two-level interactive action recognition model based on inertial data fusion. *Inform Sci* 633:264–279
- Del Din S, Godfrey A, Mazza C, Lord S, Rochester L. (2016) Free-living monitoring of parkinson's disease: lessons from the field. *Movement Disorders*, 31, 06
- Gautam Arvind, Panwar Madhuri, Biswas Dwaipayana, Acharyya Amit (2020) Myonet: a transfer-learning-based LRCN for lower limb movement recognition and knee joint angle prediction for remote monitoring of rehabilitation progress from semg. *IEEE J Trans Eng Health Med* 8:1–10
- Khan IU, Afzal S, Lee JW (2022) Human activity recognition via hybrid deep learning based model. *Sensors* 22(1):323
- Krawczyk Bartosz (2016) Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell* 5:04
- Rajesh KN, Dhuli R (2018) Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier. *Biomed Signal Process Control* 41:242–54
- Nahar J, Imam T, Tickle KS, Ali AS, Chen YP (2012) Computational intelligence for microarray data and biomedical image analysis for the early diagnosis of breast cancer. *Expert Syst Appl* 39(16):12371–7
- Taft LM, Evans RS, Shyu CR, Egger MJ, Chawla N, Mitchell JA, Thornton SN, Bray B, Varner M (2009) Countering imbalanced datasets to improve adverse drug event predictive models in labor and delivery. *J Biomed Inform* 42(2):356–364
- He H, Bai Y, Garcia E, Li SA (2008) Adasyn: Adaptive synthetic sampling approach for imbalanced learning. pp 1322 – 1328, 07
- Nur Ghaniaviyanto Ramadhan (2021) Comparative analysis of Adasyn-Svm and smote-Svm methods on the detection of type 2 diabetes mellitus. *Sci J Inform* 8(2):276–282
- Jia Hairui, Chen Shuwei (2020) Integrated data and knowledge driven methodology for human activity recognition. *Inform Sci* 536:409–430
- Sekaran SR, Han PY, Yin OS (2023) Smartphone-based human activity recognition using lightweight multiheaded temporal convolutional network. *Expert Syst Appl* 227:120132
- Han Chaolei, Zhang Lei, Tang Yin, Huang Wenbo, Min Fuhong, He Jun (2022) Human activity recognition using wearable sensors by heterogeneous convolutional neural networks. *Expert Syst Appl* 198:116764
- Roetenberg Daniel, Luinge Henk, Slycke Per (2009) Xsens mvn: full 6dof human motion tracking using miniature inertial sensors. *Xsens Motion Technol BV Tech Rep* 3:01
- Naik GR, Selvan SE, Arjunan SP, Acharyya A, Kumar DK, Ramanujam A, Nguyen HT (2018) An ICA-EBM-based sEMG classifier for recognizing lower limb movements in individuals with and without knee pathology. *IEEE Trans Neural Syst Rehabil Eng* 26(3):675–86

25. Zhang Y, Xu P, Li P, Duan K, Wen Y, Yang Q, Zhang T, Yao D (2017) Noise-assisted multivariate empirical mode decomposition for multichannel emg signals. *Biomed Eng Online* 16(1):107
26. Wang Xingjian, Dong Dengpeng, Chi Xiaokai, Wang Shaoping, Miao Yinan, An Mailing, Gavrilov Alexander I (2021) SEMG-based consecutive estimation of human lower limb movement by using multi-branch neural network. *Biomed Signal Process Control* 68:102781
27. Bansal H, Chinagundi B, Rana PS, Kumar N (2022) An ensemble machine learning technique for detection of abnormalities in knee movement sustainability. *Sustainability* 14(20):13464
28. Cai Shibo, Chen Dipei, Fan Bingfei, Mingyu Du, Bao Guanjun, Li Gang (2023) Gait phases recognition based on lower limb SEMG signals using lda-pso-lstm algorithm. *Biomed Signal Process Control* 80:104272
29. Juan Tu, Dai ZunXiang, Zhao Xiang, Huang Zijuan (2023) Lower limb motion recognition based on surface electromyography. *Biomed Signal Process Control* 81:104443
30. Venkatachalam K, Yang Z, Trojovský P, Bacanin N, Devci M, Ding W (2023) Bimodal HAR-An efficient approach to human activity analysis and recognition using bimodal hybrid classifiers. *Inform Sci* 628:542–57
31. Sola J, Sevilla J (1997) Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans Nucl Sci* 44(3):1464–8
32. Demšar Janez (2006) Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res* 7(1):1–30
33. Olive Jean Dunn (1961) Multiple comparisons among means. *J Am Stat Assoc* 56:52–64
34. Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics* 1:196–202
35. Hayashi M, Koga S, Kitagawa T (2023) Effectiveness of rehabilitation for knee osteoarthritis associated with isolated meniscus injury—a scoping review. *Cureus* 14(e34544):02
36. Wren TA, Do KP, Rethlefsen SA, Healy B (2006) Cross-correlation as a method for comparing dynamic electromyography signals during gait. *J Biomech* 39(14):2714–8
37. Al-Ayyad M, Owida HA, De Fazio R, Al-Naami B, Visconti P (2023) Electromyography monitoring systems in rehabilitation: a review of clinical applications, wearable devices and signal acquisition methodologies. *Electronics* 12(7):1520
38. Herrera-Gonzalez M, Martinez-Hernandez GA, Rodriguez-Sotelo JL, Aviles-Sanchez OF (2015) Knee functional state classification using surface electromyographic and goniometric signals by means of artificial neural networks. *Ing Univ* 1:51–66
39. Ertugrul OF, Kaya Y, Tekin R (2016) A novel approach for SEMG signal classification with adaptive local binary patterns. *Med Biol Eng Comput* 54:1137–46

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.