



A deep learning approach for parkinson's disease severity assessment

Tunç Aşuroğlu¹ · Hasan Oğul²

Received: 2 July 2022 / Accepted: 20 August 2022 / Published online: 26 August 2022

© The Author(s) 2022

Abstract

Purpose Parkinson's Disease comes on top among neurodegenerative diseases affecting 10 million worldwide. To detect Parkinson's Disease in a prior state, gait analysis is an effective choice. However, monitoring of Parkinson's Disease using gait analysis is time consuming and exhaustive for patients and physicians. To assess severity of symptoms, a rating scale called Unified Parkinson's Disease Rating Scale is used. It determines mild and severe cases. Today, Parkinson's Disease severity assessment is made in gait laboratories and by manual examination. These are time consuming and it is costly for health institutions to build and maintain laboratories. By using low-cost wearables and an effective model, aforementioned problems can be solved.

Methods We provide a computerized solution for quantifiable assessment of Parkinson's Disease symptoms severity. By using wearable sensors, our framework can predict exact symptom values to assess Parkinson's Disease severity. We propose a deep learning approach that utilizes Ground Reaction Force sensors. From sensor signals, features are extracted and fed to a hybrid deep learning model. This model is the combination of Convolutional Neural Networks and Locally Weighted Random Forest.

Results Proposed framework achieved 0.897, 3.009, 4.556 in terms of Correlation Coefficient, Mean Absolute Error and Root Mean Square Error, respectively. Proposed framework outperformed other machine and deep learning models. We also evaluated classification performance for disease detection. We outperformed most of the previous studies, achieving 99.5% accuracy, 98.7% sensitivity and 99.1% specificity.

Conclusion This is the first study to use a deep learning regression approach to predict exact symptom value of Parkinson's Disease patients. Results show that this approach can be effectively employed as a disease severity assessment tool using wearable sensors.

Keywords Gait recognition · Deep learning · Remote monitoring · Wearable sensors · Parkinson's Disease

1 Introduction

First mentioned by James Parkinson, Parkinson's Disease (PD) is a neurodegenerative disorder [1]. It is believed that in the future, PD will be the second most prevalent neurological disease after Alzheimer's disease. According to the United States Parkinson's Disease foundation, there are at least 10 million people affected by PD worldwide [2, 3]. Age is the

ultimate risk factor for developing PD. The reason behind the disease is the loss of dopamine levels in substantia nigra region of the brain and dopamine level diminishment in striatum [4]. These dopamine level diminishments lead to loss of dopaminergic neurons in the substantia nigra region [5]. Hence, medical treatments on PD patients aim to stimulate these dopaminergic neurons [6]. Unfortunately, there is no cure for PD. Furthermore, available treatments reduce disease motor symptoms just for a short while. These motor symptoms are tremor, rigidity, bradykinesia, freezing of gait, micrographia [7]. Because of these aforementioned symptoms patients become vulnerable to high-risk injuries or inability to move which leads to other health complications [8]. PD can get progressively worse as brain tissue damage increases. For this reason, since there is no permanent treatment, treatments and drugs aim to reduce the speed of PD progression [9].

✉ Tunç Aşuroğlu
tunc.asuroglu@tuni.fi

¹ Faculty of Medicine and Health Technology, Tampere University, Korkeakoulunkatu 7, P.O. Box 33720, Tampere, Finland

² Faculty of Computer Sciences, Østfold University College, P.O. Box 700, 1757 Halden, Norway

Medical treatments aim to increase and maintain the life quality of patients. Assessment of motor impairments of PD patients is done via several neurological examinations at doctor's office or at home using home diaries that is written by patients or caregivers [10]. But these short-term assessments cannot address enough and home diaries can be subjective due to its nature. Therefore, medical societies came up with several rating scales to efficiently address the severity of motor symptom impairments [10]. Most popular and extensive one is Unified Parkinson's Disease Rating Scale (UPDRS). UPDRS is a rating tool to assess the progression of PD in patients. It has benefits in distinguishing the presence of postural instability, mild and severe Parkinson's disease, and determining the life quality in moderate and severe Parkinson's disease [11]

In literature, tremor and gait abnormalities are reported as prior symptoms of PD [12]. So, in order to detect PD in a prior state, gait movement analysis is an effective choice. Because of gait movements include periodic and rhythmic foot movement patterns, these patterns emerge as an important biomarkers that not only to determine the presence of PD, but also measuring disease severity and progression [13]. Performing gait analysis is a tiresome, costly process. The analysis is conducted on hospitals or health centers that have gait laboratory infrastructure. This infrastructure mainly consists of a multi camera motion capture system to track limb movements, multiple force plates replaced over laboratory floor and muscle activity recording system [14]. With the development of wearable system technology, gait analysis has become a popular tool to assess and detect presence of PD [15]. Because the need for lab infrastructure and professional assistance can be reduced to minimum using wearable gait sensors. Ground Reaction Force (GRF) sensors are the most common sensor type for PD assessment [10]. GRF sensors can map joint movements and muscle activities effectively and it can outline characteristics of abnormal gait behavior using wearable pad sensors [15]. These gait sensors have many advantages like small size, noninvasive nature and low cost which makes them the most popular in gait analysis studies. With the advancements in machine learning, automated gait analysis can reduce time and work force problems of traditional gait analysis technologies. Also, it can solve PD severity assessment which is time consuming and exhaustive for patients and medical staff. So, with wearable sensors and machine learning, PD patients and medical staff can monitor PD symptoms in a noninvasive way with quickly and effectively.

Our motivation on this study is to provide a prognosis solution for PD patients by using easily acquirable wearable sensor data. To this end, we opt to provide a computerized solution for quantifiable assessment of PD symptoms and disease severity. By using wearable GRF sensors, our framework can forecast exact UPDRS values to assess PD

severity. We propose a multistage deep learning approach for this manner. For the first stage, several frequency and time domain features are extracted from GRF signals. After that, we utilized Convolutional Neural Networks (CNN) deep learning architecture and combine them to a Locally Weighted Random Forest (LWRF) architecture to predict UPDRS values. LWRF architecture is a local weighted Random Forest approach to reduce interpatient variability in GRF signals [10]. Our regression-based approach has two main contributions. First, by using UPDRS values our approach aims to prognose PD using a deep learning model. In our knowledge, our approach is the first study that uses deep learning regression architectures to predict exact values of PD symptom severities. Second, with our model we outperformed the previous study that use LWRF model to predict UPDRS values. In addition, we evaluated classification performance of our framework for detection of PD. To this end, several experiments are conducted to benchmark our model with previous studies on predicting PD.

The paper is structured as follows; Sect. 2 explains several related works on the same domain, Sect. 3 explains feature extraction mechanisms, our proposed deep learning approach and also dataset description. Section 4 summarizes our experimental setup and our findings. Section 5 explains advantages and disadvantages of our proposed approach.

2 Related work

Although there are many studies that use gait signals to analyze PD, these studies only focus on diagnostic conclusions about disease presence. Many studies employed traditional machine learning methods, but in recent years deep learning architectures emerge to provide further insight on disease characteristics.

Aşuroğlu et al. [10] conducted a prior study on the same problem. It is the first and only study that focuses on predicting PD symptom severity instead of binary diagnosis decisions. In this study they proposed a Locally Weighted Random Forest (LWRF) to exploit relationships between gait signals. Time and frequency features are extracted from gait signals to feed LWRF model. They reported Correlation Coefficient (CC), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values of symptom severity predictions and outperform previous studies on binary classification schemes. Lee and Lim [16] analyzed gait patterns to detect PD. In order to classify patients, they used wavelet functions to extract features from gait patterns. A fuzzy neural network is employed to conduct experiments. Daliri [17] used frequency domain features to extract meaningful information from gait signals. SVM (Support Vector Machine) model is selected as base classifier in experiments. Ertuğrul et al. [18] extracted one dimensional Local Binary Pattern

(LBP) features to classify gait signals. In their experiments they benchmarked several classifier models and features. They achieved highest accuracy with Multi-Layer Perceptron (MLP) model. Acici et al. [19] studied Random Forest (RF) architecture on classification of PD patients. Time and frequency features are extracted from multiple GRF sensors. Their approach outperformed previous studies based on classification accuracy. Zeng et al. [20] created a multi-phase classifier based on gait signals to predict PD. Feature extraction phase involves raw features such as sensor value differences between left and right foot and aggregation of left / right foot sensor output. These feature vectors are fed to a radial based neural network.

Zhao et al. [21] proposed a deep learning architecture consisting of CNN and LSTM (Long short-term memory) models. They are aimed to diagnose PD using gait signals. They reshaped gait signals in order to feed properly to CNN architecture. In their experiments, they compared their proposed model with other machine learning models and studies. In terms of accuracy, they outperformed other previous studies and machine learning models. Abdulhay et al. [22] investigated PD and tremor characteristics with gait analysis and machine learning. They extracted several features using peak detection and pulse duration measurement methods. They compared Decision Tree and SVM models for classification of PD patients. Noella et al. [23] proposed PARAFAC (Parallel Factor Analysis) and Tucker Decomposition algorithm to detect PD from multi-channel gait data. In their studies they compared their proposed model with Wavelet Transform-Multilayer Perceptron Neural Network and Hidden Markov Model-Gaussian Mixture. Their proposed framework outperformed aforementioned frameworks in terms of accuracy.

Ghaderyan and Fathi [8] proposed a new feature extraction procedure for classifying PD using gait signals. They employed singular value decomposition (SVD) model for extraction of features from multi-channel gait data. They investigated inter-limb time varying singular value (ITSV) and point out that this could be a biomarker for PD. They investigated effects of different feature spaces and sparse non-negative least-squares (NNLS) parameters on classification performance. They also investigated different PD stage classification schemes. Veeraragavan et al. [24] studied PD prediction with gait signals. They extracted multiple features including gait and statistical features. They fed these extracted features to an Artificial Neural Network (ANN) classification model. They compared performance of different cross validation and minority sampling methods. They also reported PD stage classification results for their proposed model. Xia et al. [9] proposed a deep learning-based approach for classifying PD using gait sensor data. They combined CNN and LSTM architectures to extract local

deep features from gait signals. They compared different feature extraction schemes and machine learning models in their experiments. Their approach outperformed previous studies and other models in terms of accuracy, sensitivity and specificity.

Priya et al. [25] explored effects of different local binary pattern techniques on PD diagnosis. They transformed gait data using LBP algorithms. After transformation they extracted statistical features from these patterns. In their experiments, they compared classifier performances of KNN, SVM and Logistic Regression models. They reported that Logistic regression model with LBP patterns achieved highest accuracy, precision and recall. Liu et al. [26] conducted studies on PD diagnosis using a dual branch deep learning model. In this model they combined CNN and Bi-LSTM architectures for each foot's gait signal. They extracted features from each gait cycle. Comparison of their proposed model with other models such as KNN, RF and CNN stated that their method outperformed others in terms of accuracy, sensitivity and specificity. Tong et al. [15] conducted studies using permutation-variable importance and persistent entropy to classify different PD severity levels. They used aforementioned feature extraction scheme with RF to get gait features. After that they used SVM model to classify these gait features. A summarization of related works that includes features and classifiers are given in Table 1.

3 Materials & methods

Gait signal processing is done using multiple gait sensors that measures GRF value. GRF value is a force that floor applies when walking. This force can have different values for different walking activities and thus it is a popular tool for personal gait analysis research [17]. Our proposed machine learning framework takes multi-channel GRF signal values from wearable sensors and utilize a deep regressor model for finding relationships between walking patterns and motor symptom severities. We employed a simple learning model that includes training and testing. Firstly, frequency and time domain features of annotated sensor signals are constructed from dataset. In our framework, constructed feature vectors depict PD or healthy participants in the dataset. Secondly, a deep regressor/classifier is trained using these feature vectors. Every signal sample has an exact UPDRS scale value and a class label (PD, Healthy). Thirdly, regressor/classifier predicts symptom severity or class value of a test sample. As final step, evaluation of trained model is done via several metrics. Workflow of proposed approach is summarized in Fig. 1.

Table 1 Features and classifier models of previous studies on PD classification

Authors	Feature	Classifier
Lee and Lim [16]	Wavelet based features	A fuzzy neural network
Daliri [17]	Frequency features	SVM
Ertuğrul et al. [18]	LBP	MLP
Açici et al. [19]	Time and frequency features	RF
Aşuroğlu et al. [10]	Time and frequency features	LWRF
Zeng et al. [20]	Gait raw data	Radial based neural network
Zhao et al. [21]	Gait raw data	Combination of CNN and LSTM
Abdulhay et al. [22]	Kinetic and temporal features	SVM
Noella et al. [23]	Gait raw data	PARAFAC and Tucker decomposition algorithm
Ghaderyan and Fathi [8]	Inter-limb time-varying singular value	sparse non-negative least-squares
Veeraragavan et al. [24]	Gait and statistical features	ANN
Xia et al. [9]	Gait raw data	Combination of CNN and LSTM
Priya et al. [25]	LBP	Logistic regression
Liu et al. [26]	Gait raw data	Combination of CNN and LSTM
Tong et al. [15]	Gait characteristic features	SVM

3.1 Feature extraction

It is not suitable to feed GRF signals to our regression framework directly. First of all, many regressor models require equal input length, which is not the case for GRF signals. GRF signals have varying input lengths. Second, GRF signals are not in the same time scale even if they have equal length. Hence, finding similar local patterns is hard. Third, temporal characteristics sometimes cannot be able to reflect

motion of the signal in terms of UPDRS values that the regressor is asked to find. For this reason, when analyzing such time series signal data, a common approach is to construct a feature vector from raw signal samples.

We applied same feature extraction procedure from a previous study [10] in order to establish the same experimental setup for data analysis. For this end, two type of features are selected to represent raw signal values: time and frequency domain features. Sixteen time and seven

Fig. 1 Workflow of proposed approach

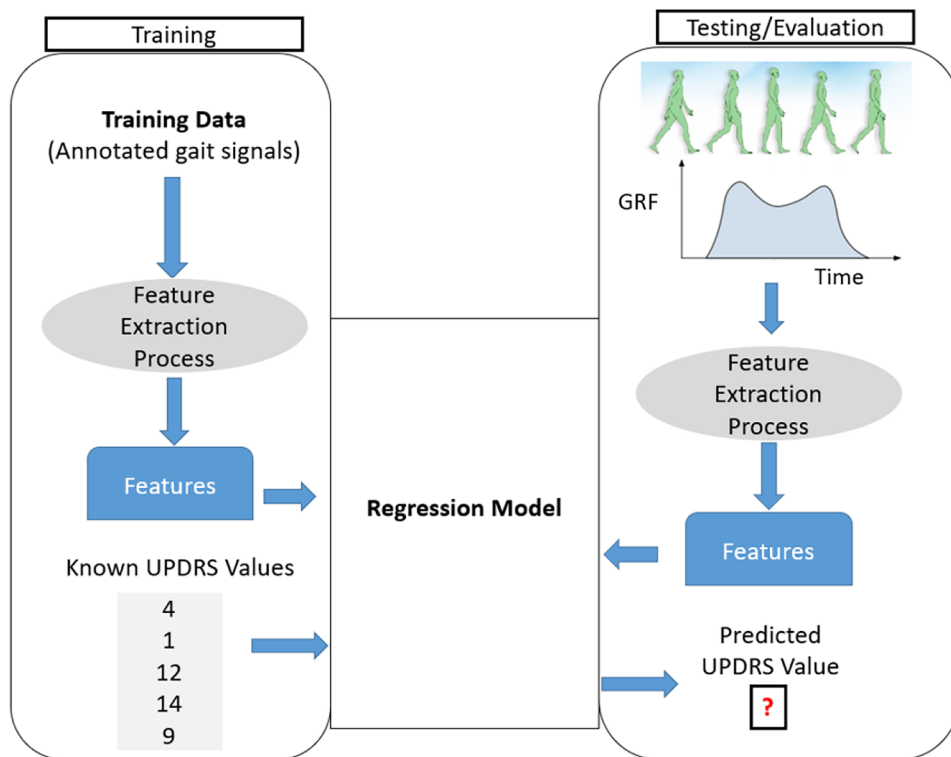


Table 2 Extracted features from each sample in dataset

Feature Domain	Features
Frequency	mean, minimum and maximum value normalized value, energy, power and phase
Time	mean, minimum/maximum value and their index, range, mean absolute deviation, median, interquartile range (IQR), harmonic mean, kurtosis, root mean square (RMS), energy, power and entropy, skewness

frequency domain features are extracted for each GRF signal [10, 27]. In the literature, time and frequency features of a time series signal is proved to bring out important characteristics of a signal [10, 17, 27–29]. Extracted features are given in Table 2.

3.2 Dataset

For this study, we used “Physionet Gait in Parkinson's Disease” public access dataset that is comprised of gait data [30]. The dataset consists of gait signal measurements for a control group and PD patients. PD patients average age is 66.3 and 37% of patients are women. Control group average age is 63.7 years and 45% of control group are women. There are 93 PD patients and 73 control subjects are included in experiments. Gait signal measurement tests were conducted during walking for 2 min. In order to record gait signal measurements dual-task and usual walking procedures are applied [31]. In usual walking procedure, subjects were asked to walk at their normal walking speed on a level ground for 20 m and go back to their starting location. In dual tasking procedure, subjects were asked to walk the same path at their normal walking speed while being asked to continuously extract number 7 from a pre-defined number (e.g., 300, 293, 286, 279) [31].

Gait phase inspection before and post rehabilitation, provides a potent treatment. Gait signal stages can be defined by pressure measurements because foot pressure distributions are repetitive and unique for each phase [32]. Force sensitive sensors (FSR) are often employed to evaluate these distributions. The main drawback of these systems is their limited detection area and reach. For this reason, FSR sensors do not reflect accurate foot pressure measurements. GRF measuring wearable smart shoes become popular due to negative sides of FSR sensors [33].

To measure gait signals, eight sensors are appended to each foot. Force value (Newton) for each sensor is measured as a function of time. GRF sensor positions can be seen in Fig. 2.

All GRF sensor measurements have a 100 Hz sampling rate. In order to reduce starting and ending effect of measurements, 20 s from beginning and 10 s from end were deleted. To remove outlier values, a median filter was applied. This preprocess step diminishes the fluctuations of gait motion, therefore achieving an accurate representation of gait characteristic for every subject.

Gait dataset also has several demographic data for each subject. Also, it has severity of PD values as Unified Parkinson's Disease Rating Scale (UPDRS), a measure that is used to monitor quantitative symptoms. UPDRS is widely used as an evaluation tool for patients. It is made of several parts: (a) motor examination; (b) daily living activities (c) the complications of therapy; (d) mentation, behavior, and mood [11]. UPDRS has values 0 to 199. UPDRS value distribution of PD patient data cohort is as follows; minimum UPDRS value is 13, average is 32 and maximum is 70.

3.3 Proposed deep learning approach

Our proposed framework is a hybrid deep learning approach that consists of CNN and LWRF architectures to utilize data relations. Our main objective consists of extracting local CNN features from gait signals with convolutional layers and combine them with LWRF architecture to find local relations between features. LWRF model is a locally weighted Random Forest

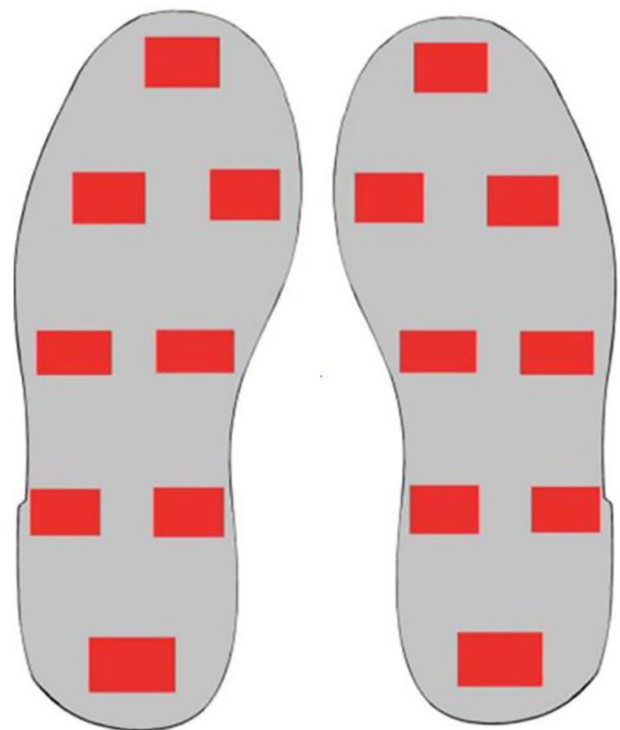


Fig. 2 GRF sensor placements for each foot

approach that utilizes local relationships between neighboring data [10]. With the combination of machine and deep learning models, data relationships can be mapped more efficiently [34].

Random Forest is a member of ensemble learning models. Random Forest does ensemble learning using an approach called Bootstrap Aggregation (Bagging). In Bagging, a random bootstrap sample is used to construct decision trees [19]. Then, model forms a collection of decision trees with training dataset variables. These trees are called Random Trees. While forming new branches in a tree, algorithm selects variables randomly instead of exploring best branches in all of the variables. This approach helps reducing correlation between random trees. If there is a high correlation between trees, trees can converge to same output and thus reduce prediction accuracy. Random tree collections produce class or numeric values depending on machine learning task. Last step in Random Forest algorithm is majority voting process on decisions.

LWRF depends on a learning process that is locally weighted [35]. LWRF model exploits local relationships that only utilize sub datasets. In locally weighted learning process, when creating a model, a local model is constructed from nearest to the query point instead of building a globalized model for functional space. Every data point has assigned a weighted value. Weight values influence predictor approach when estimating a target value [36]. Closest data points to query have bigger weighted values comparing with far points. Adaptation of this approach is applied to Random Forest model via locally weight each data point in computing split points and selecting bootstrap samples for decision trees [10].

Although CNN-based deep learning models are often used in image and video domains [37, 38], in recent years several applications of them have emerged in other signal processing studies [39, 40]. CNN layers expose local relationships in data. CNN neurons split a given image to many local regions and construct features from these regions. A CNN architecture consists of a layer of input, a layer of output and numerous hidden layers [40]. Convolutional layers can be transformed to a feature extractor or a classification/regression tool. Our proposed model employs CNN model as tool for deep feature extraction. We used CNN to extract deep local features of stacked sensor signal channels. The reason behind this is to utilize CNN as a deep local relationship feature extractor instead of just using hand crafted signal features.

By constructing this hybrid deep learning model with deep local features and LWRF, our approach aims to capture the important descriptors of GRF signals using Convolutional layers, represent them by collection of random trees in a locally weighted approach that utilizes valuable local descriptors that represent data.

All gait samples in the dataset have one dimension after time and frequency features are extracted. Each sample has a total length of 368 (23 features \times 16 sensors). In our proposed approach, we first convert these gait samples to

multi-channel samples so that we can process them with CNN architecture. Each gait sample is converted to a sixteen-channel sample. Constructed multichannel samples have 23 values along x axis, which represents seven frequency and sixteen time-domain features and have 16 channels which corresponds to gait sensors on each foot. Multi-channel conversion of gait samples is shown in Fig. 3.

After all samples are converted to multi-channel samples, they are ready to be fed to a hybrid model. This hybrid model is made of a layer of input, convolution layer blocks and LWRF algorithm to predict UPDRS values of extracted local deep features. Our proposed model can be seen in Fig. 4. In CNN part of our model, there are four convolution layers which have 40 filters. These filters vary in filter sizes which are 1×6 and 7×1 . To extract vertical and horizontal local features of gait signals we selected these filter sizes. Horizontal local features exploit relationship between time and frequency domain features whereas vertical local features exploit relationship between multiple GRF sensors. Therefore, combination of these valuable local features could ensure a better representation than just hand-crafted time and frequency domain features. A normalization layer, a max pooling and a ReLU layer is added in CNN architecture. Deep features that are extracted from last convolution layer are fed as an input for LWRF model.

4 Results

4.1 Implementation

MathWorks MATLAB 2020a tool is selected for model implementation. GeForce CUDA library is employed as a booster for deep learning architecture speed. Experiments implemented on a computer with a computer that has 16 GBs of RAM, Intel Core i7 2.2 GHz Processor and GeForce GTX 1060 graphics card.

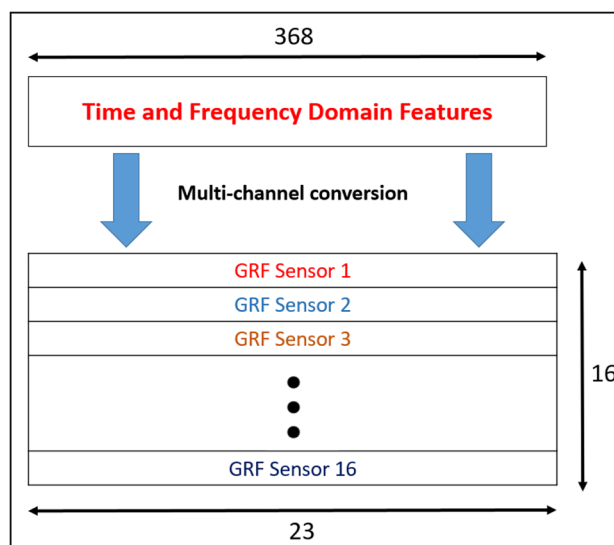
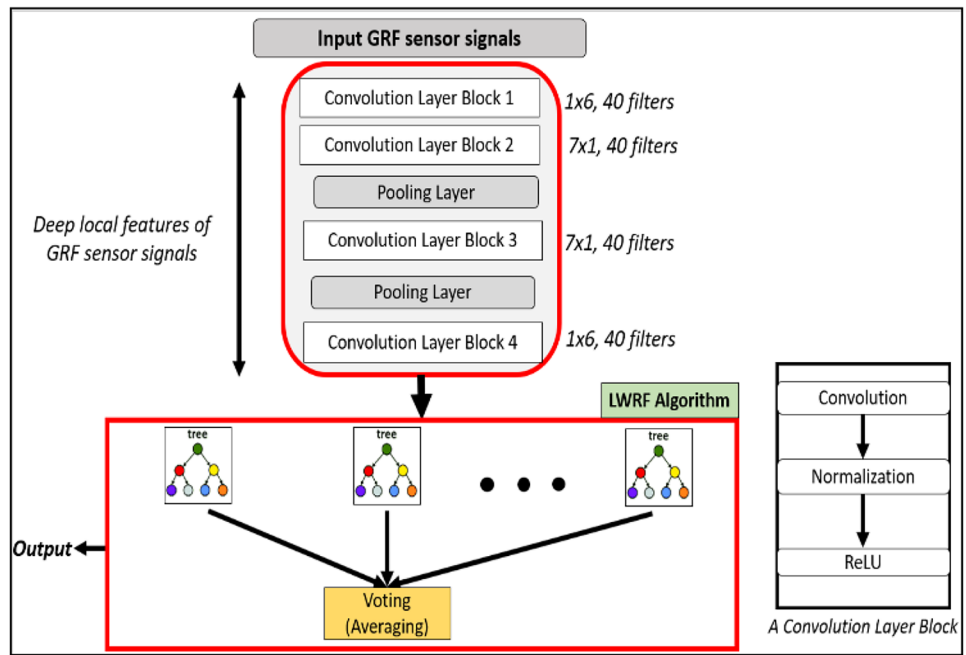


Fig. 3 Conversion process of GRF signals

Fig. 4 Proposed model for PD severity prediction



4.2 Evaluation

Experiments are conducted using a tenfold cross validation (CV) approach. For tenfold CV approach, dataset is split to ten parts and then a single part is reserved for testing while other parts are act as training set. CV stops to function when testing phase is done with all parts.

In order to evaluate our hybrid deep learning model performance, we used Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Correlation Coefficient (CC)). In order to achieve good performance, our model should have low error rate in terms of MAE and RMSE and also high CC value.

Calculation of CC value depends on several variables, n corresponding to size of sample, p_i and a_i are estimated and real values, \bar{p} and \bar{a} are mean values. Relationship between estimated and real values are calculated via CC value. It can have values that range between 1 and -1. Full negative correlation is represented as -1 while full positive correlation is represented as +1. Formulation of CC:

$$CC = \frac{S_{PA}}{\sqrt{S_P S_A}} \tag{1}$$

$$S_{PA} = \frac{\sum_i (p_i - \bar{p})(a_i - \bar{a})}{n - 1},$$

$$S_P = \frac{\sum_i (p_i - \bar{p})^2}{n - 1}, \text{ and}$$

$$S_A = \frac{\sum_i (a_i - \bar{a})^2}{n - 1}.$$

MAE is a measure of the difference between two continuous variable:

$$MAE = \frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n} \tag{2}$$

RMSE is a quadratic metric that measures the magnitude of error of a machine learning model, which is often used to find the distance between the predicted values of the estimator and the real values.:

$$RMSE = \sqrt{\frac{(|p_1 - a_1|)^2 + \dots + (|p_n - a_n|)^2}{n}} \tag{3}$$

To study our model’s classification capability, we used well known classifier evaluation metrics: Accuracy (Acc.), Sensitivity (Sn.), Specificity (Sp.) and F-Score (4). TP, TN, FP, FN correspond to True Positive, True Negative, False Positive and False Negative respectively.

$$Acc = \frac{TP + TN}{(TP + TN + FP + FN)}, Sn = \frac{TP}{TP + FN},$$

$$Sp = \frac{TN}{TN + FP}, F - Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{4}$$

Hyper-parameter settings of deep learning model are implemented by a tuning approach. Optimization of parameters are evaluated by error for predictions. We first begin by roughly estimate of parameters aiming to achieve lowest error and highest correlation. We then select hyper-parameters for our deep learning model that give lowest prediction error in experiments. Epoch is selected as 50

and Adam optimization algorithm selected for training the model. Same value padding is done in convolution layers. For normalization layer, local response normalization for each channel is employed. Loss function is selected as Mean squared error for deep learning model.

4.3 Empirical results

We employed two basic comparison approaches in experiments, one is for regression setup and the other one is for classification setup. Since our main goal is to assess model performance on predicting real symptom values, regression based experimental setup is used. The reason for selecting classification setup is to further assess our model's ability to detect PD. In order to compare our model's performance, we conducted experiments on several selected models. Selected deep learning models are LSTM [41], CNN [40], CNN + LSTM [42, 43] and CNN + RF [44, 45] architectures. Selected machine learning models are LWRF [10], Random Forest (RF) [46], Support Vector Regression (SVR) [47], J48 decision tree [48], Linear regression (LIR), Regression by Discretization (RBD) [49] and k-NN [50]. Selected CNN model for experiments consists of four convolutional layer blocks (Convolution, Normalization, ReLu layers), two pooling layers and a regression output layer. This model uses the same hyper-parameter setup as proposed model. 128 neuron Bi-LSTM architecture is selected as LSTM architecture and He optimizer selected as weight initializer and soft sign is assigned as activation function. For combined deep learning models, we employed same CNN model architecture for deep feature extraction.

Experimental results for regression models to predict UPDRS values is given in Table 3. Proposed deep learning model achieved 0.897, 3.009, 4.556 in terms of Correlation Coefficient (CC), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), respectively. Our model comes on top among other regressor models considering all metrics. CNN architecture comes as second. Linear regression

Table 3 UPDRS value prediction performance of regression models

Model Name	CC	MAE	RMSE
k-NN (k=1)	0.873	4.703	8.113
k-NN (k=27)	0.765	8.564	10.896
RF	0.876	5.678	8.124
RBD (RF)	0.878	5.763	8.072
LWRF [10]	0.895	4.462	7.382
SVR	0.867	5.308	8.298
J48	0.754	6.609	11.439
LIR	0.679	11.649	15.194
CNN	0.884	3.267	4.689
LSTM	0.624	10.371	13.303
CNN + LSTM	0.730	8.72	10.31
CNN + RF	0.874	4.577	5.758
Proposed Model	0.897	3.009	4.556

and LSTM algorithms performed poorly for predicting UPDRS values.

Another conclusion is that our hybrid model outperformed base model (LWRF) considering all of the metrics. Also, CNN architecture performed better than LWRF. Our experiments point out that learned deep features have more representative information than hand crafted ones when predicting UPDRS value of samples. We can come to a conclusion that LWRF model can handle CNN features very well for PD symptom monitoring.

LWRF learning model considers one parameter when making a prediction. It is the number of nearest neighbors (k). k value is selected as 27 in experiments. The reason for this selection can be seen in Fig. 5, highest recorded CC value is around 27 and decreases after this value.

Although our approach looks through a regression window, we also wanted to validate our approach from a classification perspective. To this end, a classification experiment setup was created and our proposed model were compared with previous PD detection studies. For this purpose, our

Fig. 5 Effect of k value on UPDRS prediction

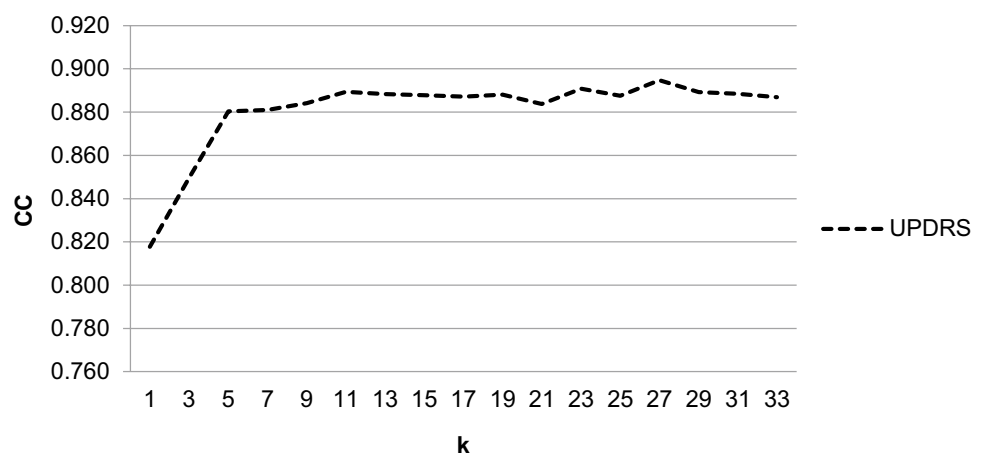


Table 4 Comparison with previous studies on PD classification

Authors	Acc. (%)	Sn. (%)	Sp. (%)	F Score
Lee and Lim [16]	74.3	81.6	73.8	NR
Daliri [17]	91.2	91.71	89.92	NR
Ertuğrul et al. [18]	88.8	88.9	82.2	NR
Açici et al. [19]	98	99.1	95.7	0.98
Aşuroğlu et al. [10]	99	97.8	99.5	NR
Zeng et al. [20]	98.80	98.92	98.63	NR
Zhao et al. [21]	98.61	NR	NR	NR
Abdulhay et al. [22]	94.8	NR	NR	NR
Noella et al. [23]	97.0	NR	NR	NR
Ghaderyan and Fathi [8]	97.22	98.22	95.86	NR
Veeraragavan et al. [24]	97.7	97.05	97.41	0.97
Xia et al. [9]	99.07	99.1	99.01	NR
Priya et al. [25]	98.82	NR	NR	NR
Liu et al. [26]	99.22	98.04	100	0.99
Tong et al. [15]	99.23	NR	NR	NR
Present method	99.5	98.7	99.1	0.99

model was transformed in a way to do PD diagnosis. Same feature set and data samples from regression setup are used in classification scheme, but sample labels are replaced with PD class (0 for control, 1 for disease) instead of real UPDRS values. Table 4. shows this comparison between other studies. According to Table 4, our model which CNN attributes are fed as input to the LWRF algorithm, shows the best performance with an accuracy (Acc.) of 99.5%. Our model outperforms most of the studies in terms of Sensitivity (Sn.) and Specificity (Sp.). Finally, our model achieves a higher rate than most of the available studies in terms of F-score. (NR: shows results that have not been reported).

5 Conclusion & discussion

Our approach aims to facilitate remote monitoring of motor symptoms of PD patients by predicting UPDRS values. Monitoring PD patients remotely has an essential role for providing a satisfied clinical and home care. Gait signal measurements play another essential role for predicting PD severity because most of these patients lost their basic motor abilities such as walking. Today, assessment of PD severity is made in a specialized gait laboratories and sometimes by manual examination of clinicians. These processes are time consuming for patients and clinicians. In another perspective, it is costly for health institutions to maintain and build such sophisticated lab infrastructures. With the development of wearable system technologies gait analysis has become popular tool to assess and detect presence of PD. One of the most popular one is GRF sensor. These wearable GRF sensors can be used easily to detect and assess PD severity.

As can be seen from our results, by using wearable GRF sensors, disease progress can be continually monitored without a lab infrastructure. So, by using low-cost wearable sensors and effective machine learning model, our solution can overcome aforementioned problems. Our regression-based approach has two main contributions. First, by using UPDRS values our approach aims to prognose PD using a deep learning model. In our knowledge, our approach is the first study that uses deep learning architectures to predict exact values of PD symptom severities. Second, with our model we outperformed previous study that use LWRF model to predict UPDRS values.

A hybrid deep learning model is proposed in this study. In this hybrid deep learning model, converting time and frequency features of GRF sensors and gave as input to CNN + LWRF architecture performed better than other regressor models. More specifically, our proposed model outperformed previous study that used LWRF model in all of the evaluation metrics. Conducted experiments stated that learned deep features have more representative information than hand crafted ones when predicting UPDRS value of samples. Another conclusion from our experiments is that layers of CNN utilize relationships between multiple GRF sensors and LWRF model can learn these relationships well with its locally weighted structure. To take further our research we also modified our regressor model to classifier to detect the presence of PD. Our algorithm performed better in terms of accuracy than all of the previous studies. This result points out that our model can also be used to detect presence of PD from gait signals.

Our proposed framework has some limitations. The gait dataset that we used in our studies has limited PD patient population. This is a known issue in PD studies since it's hard to find PD patients and convince them to participate in such research. But ongoing studies aim to add more PD patient data in relevant datasets. Another solution to this problem is using resampling algorithms to increase PD sample size. ADASYN or SMOTE algorithms can be used in this manner. By resampling, performance of our deep learning framework can be increased. Because deep learning models require a lot of data to perform better in all terms. Another issue is the heterogeneous property of dataset in terms of disease severity values. Patients with higher UPDRS values usually have big problems with walking and dual tasking. Therefore, to preserve safety of these patients they are not included in the dataset [7]. This leads to higher number of low and medium UPDRS values.

In the future, we opt to compare our model's performance with different gait datasets. In addition, benchmarking our model with different neurological disorders such as Huntington disease and Alzheimer's can confirm our model's validity. Also, these comparisons can validate our model's generalization capability. In the future, we also aim to test our model with

other PD severity rating scales such as PDQ-39, MDS-UPDRS. Another future direction for this study is, utilizing hyper parameter estimation methods such as grid search and meta-heuristic approaches to further decrease prediction error. By analyzing hyper parameters thoroughly, proposed approach can increase its robustness. Finally, other available hybrid deep learning architectures [51] and deep auto-encoders [38] can be integrated into our model to increase prediction performance.

Authors' contributions All the authors contributed equally to this research. visualization, investigation, formal analysis and writing of the original draft were performed by Tunç Aşuroğlu. Conceptualization, Methodology, Editing and Review were conducted by Hasan Oğul.

Funding No funding was received for conducting this study.

Availability of data Not applicable.

Code availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflicts of interest The authors have no conflicts of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Parkinson J. An essay on the shaking palsy. *J Neuropsychiatry Clin Neurosci.* 2002;14(2):223–36.
- Lombardo JM, Lopez MA, Miron F, López M, León M, Arambarri J, Álvarez D. MOBEEZE Natural Interaction Technologies, Virtual Reality and Artificial Intelligence for Gait Disorders Analysis and Rehabilitation in Patients with Parkinson's Disease. *Int J Interact Multi Artif Intell.* 2019;5:54–62.
- Mack S. Removing Unclassified Hand Tremor Motion from Computer Mouse Input with Neural Networks. *Int J Interact Multi Artif Intell.* 2018;5:56–60.
- Xu S, Pan Z. A novel ensemble of random forest for assisting diagnosis of Parkinson's disease on small handwritten dynamics dataset. *Int J Med Inform.* 2020;144:104283, 12.
- Delva A, Van Weehaeghe D, Koole M, Van Laere K, Vandenberghe W. Loss of Presynaptic Terminal Integrity in the Substantia Nigra in Early Parkinson's Disease. *Mov Disord.* 2020;35:1977–86.
- Cabreira V, Soares-da-Silva P, Massano J. Contemporary Options for the Management of Motor Complications in Parkinson's Disease: Updated Clinical Review. *Drugs.* 2019;79:593–608.
- Silva de Lima AL, Hahn T, de Vries NM, Cohen E, Bataille L, Little MA, et al. Large-scale wearable sensor deployment in Parkinson's patients: the Parkinson@home study protocol. *JMIR Res Protoc.* 2016;5(3):e172.
- Ghaderyan P, Fathi G. Inter-limb time-varying singular value: A new gait feature for Parkinson's disease detection and stage classification. *Measurement.* 2021;177:109249.
- Xia Y, Yao Z, Ye Q, Cheng N. A dual-modal attention-enhanced deep learning network for quantification of Parkinson's disease characteristics. *IEEE Trans Neural Syst Rehabil Eng.* 2020;28(1):42–51.
- Aşuroğlu T, Açıcı K, Erdaş CB, Toprak MK, Erdem H, Oğul H. Parkinson's disease monitoring from gait analysis via foot-worn sensors. *Biocybern Biomed Eng.* 2018;38(3):760–722.
- Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Mov Disord.* 2008;23:2129–70.
- Balaji E, Brindha D, Balakrishnan R. Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease. *Appl Soft Comput.* 2020;94.
- Creaby MW, Cole MH. Gait characteristics and falls in Parkinson's disease: A systematic review and meta-analysis. *Parkinsonism Relat Disord.* 2018;57:1–8.
- Chen PH, Wang RL, Liou DJ, Shaw JS. Gait disorders in Parkinson's disease: assessment and management. *Int J Gerontol.* 2013;7(4):189–93.
- Tong J, Zhang J, Dong E, Du S. Severity Classification of Parkinson's Disease Based on Permutation-Variable Importance and Persistent Entropy. *Appl Sci.* 1834;11:2021.
- Lee S-H, Lim JS. Parkinson's disease classification using gait characteristics and wavelet-based feature extraction. *Expert Syst Appl.* 2012;39(8):7338–44.
- Daliri MR. Chi-square distance kernel of the gaits for the diagnosis of Parkinson's disease. *Biomed Signal Process Control.* 2013;8(1):66–70.
- Ertuğrul ÖF, Kaya Y, Tekin R, Almalı MN. Detection of Parkinson's disease by shifted one dimensional local binary patterns from gait. *Expert Syst Appl.* 2016;56:156–63.
- Açıcı K, Erdaş ÇB, Aşuroğlu T, Toprak MK, Erdem H, Oğul H. 'A random forest method to detect Parkinson's disease via gait analysis.' In: Boracchi G, Iliadis L, Jayne C, Likas A, editors. *Engineering Applications of Neural Networks (EANN) (Communications in Computer and Information Science)*, vol. 744. Cham, Switzerland: Springer; 2017. p. 609–19.
- Zeng W, Liu F, Wang Q, Wang Y, Ma L, Zhang Y. Parkinson's disease classification using gait analysis via deterministic learning. *Neurosci Lett.* 2016;633:268–78.
- Zhao A, Qi L, Li J, Dong J, Yu H. A hybrid spatio-temporal model for detecting and severity rating of Parkinson's disease from gait data. *Neurocomputing.* 2018;315:1–8.
- Abdulhay E, Arunkumar N, Narasimhan K, Vellaiappan E, Venkatraman V. Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease. *Futur Gener Comput Syst.* 2018;83:366–73.
- Noella N, Gupta D, Priyadarshini J. Diagnosis of Parkinson's disease using gait dynamics and images. *Proc Int Conf Recent Trends Adv Comput.* 2019.

24. Veeraragavan S, Gopalai AA, Gouwanda D, Ahmad SA. Parkinson's Disease Diagnosis and Severity Assessment Using Ground Reaction Forces and Neural Networks. *Front Physiol.* 2020;11:587057.
25. Priya SJ, Rani AJ, Ubendran N. Improving the prediction accuracy of Parkinson's Disease based on pattern techniques. In: *Proceedings 5th International Conference on Devices, Circuits and Systems (ICDCS), 2020*, pp. 188–192.
26. Liu X, et al. A dual-branch model for diagnosis of Parkinson's disease based on the independent and joint features of the left and right gait. *Appl Intell.* 2021;51:7221–32.
27. Açıcı K, Erdaş CB, Aşuroğlu T, Oğul H. Handy: A Benchmark Dataset for Context-Awareness via Wrist-Worn Motion Sensors. *Data.* 2018;3(3).
28. Erdaş ÇB, Atasoy I, Açıcı K, Oğul H. Integrating Features for Accelerometer-based Activity Recognition. *Procedia Comput Sci.* 2016;98:522–527.
29. Climent-Pérez P, Florez-Revuelta F. Privacy-Preserving Human Action Recognition with a Many-Objective Evolutionary Algorithm. *Sensors.* 2022;22(764).
30. Gait in Parkinson's Disease Dataset. 2008. [online] Available: <https://physionet.org/content/gaitpdb/1.0.0/>. Accessed 17 Aug 2022.
31. Hausdorff JM, Balash J, Giladi N. Effects of cognitive challenge on gait variability in patients with Parkinson's disease. *J Geriatr Psych Neurol.* 2003;16:53–8.
32. Bae J, Tomizuka M. Gait phase analysis based on a hidden Markov model. *J Mechatron.* 2011;21(6):961–70.
33. Bae J, Tomizuka M. A tele-monitoring system for gait rehabilitation with an inertial measurement unit and shoe-type ground reaction force sensor. *Mechatronics.* 2013;23(6):646–51.
34. Aşuroğlu T, Oğul H. A deep learning approach for sepsis monitoring via severity score estimation. *Comput Methods Programs Biomed.* 2021;198(105816).
35. Atkeson CG, Moore AW, Schaal S. Locally weighted learning. In: *Lazy learning*. Springer; 1997. pp. 11–73.
36. Englert P. Locally weighted learning. *Seminar Class Auton Learn Syst.* 2012;1(1):1–9.
37. Karim AM, Güzel MS, Tolun MR, Kaya H, Çelebi FV. A new generalized deep learning framework combining sparse autoencoder and Taguchi method for novel data classification and processing. *Math Problems Eng.* 2018;2018.
38. Karim AM, Güzel MS, Tolun MR, Kaya H, Çelebi FV. A new framework using deep auto-encoder and energy spectral density for medical waveform data classification and processing. *Biocybernetics Biomed Eng.* 2019;39(1):148–59.
39. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B. Recent advances in convolutional neural networks. *CoRR.* 2017;187–332.
40. Açıcı K, Aşuroğlu T, Erdaş ÇB, Oğul H. T4SS Effector Protein Prediction with Deep Learning. *Data.* 2019;4(45).
41. Fagerström J, Bång M, Wilhelms D, Chew MS. LiSep LSTM: A machine learning algorithm for early detection of septic shock. *Sci Rep.* 2019;9(1):1–8.
42. Tan JH, et al. Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals. *Comput Biol Med.* 2018;94:19–26.
43. Ordóñez FJ, Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors.* 2016;16(1):115.
44. Wang A, Wang Y, Chen Y. Hyperspectral image classification based on convolutional neural network and random forest. *Remote Sens Lett.* 2019;10(11):1086–94.
45. Knauer U, von Rekowski CS, Stecklina M, Krokotsch T, Minh TP, Hauffe V, et al. Tree species classification based on hybrid ensembles of a convolutional neural network (CNN) and random forest classifiers. *Remote Sens.* 2019;11(23):2788.
46. Verikas A, Gelzinis A, Bacauskiene M. Mining data with random forests: A survey and results of new tests. *Pattern Recognit.* 2011;44(2):330–49.
47. Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK. Improvements to the SMO algorithm for SVM regression. *IEEE Trans Neural Netw.* 2000;11(5):1188–93.
48. Acharya TD, Lee DH, Yang IT, Lee JK. Identification of water bodies in a Landsat 8 OLI image using a J48 decision tree. *Sensors.* 2016;16(7).
49. Salman R, Kecman V. Regression as classification. *Southeastcon Proc IEEE.* 2012.
50. Amrane M, Oukid S, Gagaoua I, Ensarî T, “Breast cancer classification using machine learning,”. *Electric Electronics, Computer Science. Biomedical Engineerings' Meeting (EBBT).* 2018;2018:1–4.
51. Sivari E, Güzel MS, Bostanci E, Mishra A. A Novel Hybrid Machine Learning Based System to Classify Shoulder Implant Manufacturers. *Healthcare.* 2022;10(580).

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