



Parkinson's disease diagnosis and stage prediction based on gait signal analysis using EMD and CNN-LSTM network

B. Vidya, Sasikumar P.*

School of Electronics Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, 632014, India



ARTICLE INFO

Keywords:

Gait analysis
Parkinson's disease
VGRF
EMD
CNN-LSTM

ABSTRACT

Parkinson's disease (PD) diagnosis is a complex and challenging task which needs the assessment of various motor and non-motor symptoms. As gait impairment is one of the early and important symptoms of PD, in a clinical setting, physicians generally evaluate the gait abnormality based on visual observations along with other numerous manifestations to assess the severity of PD. As such kind of assessment majorly depends on the experience and expertise of the physicians, there is scope for bias in assessment, leading to misdiagnosis. In this context, to assist the physicians to diagnose PD effectively, this study aims to design and investigate a gait analysis based classifier model using a hybrid convolutional neural network-long short term memory (CNN-LSTM) network to predict the severity rating of PD. For evaluation, we utilize the openly available gait dataset from Physionet that consists of vertical ground reaction force (VGRF) signals from three different walking tests. Firstly, the prominent VGRF signals obtained using the variability analysis are decomposed using the empirical mode decomposition (EMD) technique to extract the significant intrinsic mode functions (IMFs) that contain the vital gait features. Secondly, through the power spectral analysis the dominant IMFs of the selected VGRF signals are extracted to train the CNN-LSTM classifier model. To address the data overfitting problem in the classifier model, the proposed approach employs L2 regularization along with dropout techniques. Moreover, to solve the stochastic cost function, CNN-LSTM network utilizes the Adam optimizer for its minimal memory requirement and tuning. Finally, the experiments conducted using the gait patterns from 93 PD subjects and 73 healthy controls substantiate that the proposed CNN-LSTM classifier model can achieve a maximum multi-class classification accuracy of 98.32% and offer superior performance compared to several other similar methods which have used gait pattern to diagnose PD.

1. Introduction

Parkinson's disease is a idiopathic, chronic neurodegenerative disease which affects not only the regulation of movements but also emotions due to the deficit of a neurotransmitter in the brain called "dopamine". The statistics from the Parkinson's foundation, USA, reveals that approximately 12 million people across the globe are diagnosed annually with PD (Mei et al., 2021). Even though the causes of PD are still largely unknown, the early diagnosis can help to identify suitable medication/therapy to delay full onset of PD symptoms. According to the severity, nearly half of PD patients manifest several dysfunctions including gait abnormality, attention deficit, speech problems and impulsivity (Camps et al., 2018; Goyal et al., 2020). The cardinal PD motor symptoms are tremor, gait imbalance, postural instability and bradykinesia. Similarly, the primary non-motor symptoms of PD include fatigue, dementia, depression and sleep disorder (Ricciardi et al., 2019). Currently, several therapies are available for treating PD

including surgical treatments, gene therapies, pharmacological manipulations and rehabilitation therapies. As discussion on the PD therapies is not within the scope of this article, we recommend the interested readers to Maiti et al. (2017) for the detailed overview on current diagnostic strategies and therapies.

From diagnostic standpoint, the biosignals such as speech (Zhao et al., 2019), gait (Wahid et al., 2015), and handwriting (Kamran et al., 2021) have been utilized to assess the motor impairments and identify the onset of PD. Since the symptoms and variability of PD progression are heterogeneous in nature, there has been significant interest in developing decision support systems that can predict the progression of PD (Farashi, 2020). The gait analysis of the subject provides crucial information such as joint dynamics, kinematic parameters, and appropriate strategy for surgical treatment and rehabilitation. Recent investigations substantiate that gait cycle investigation can act as an efficient approach to identify the presence of PD because the postural instability is one of the early manifestations of PD (Cantürk, 2021). Some of the techniques available for gait recording are: motion cameras

* Corresponding author.

E-mail address: sasikumar.p@vit.ac.in (Sasikumar P.).

to capture the gait trajectory, inertial measurement unit (IMU) device to compute the foot angular movement, and pressure sensors to read foot plantar pressure, and force sensitive resistors (FSR) to acquire the vertical ground reaction forces.

Among the aforementioned techniques, the gait recording system using insoles with FSR is an easy, quick and cost-effective technique. In addition, the quantitative assessment of kinematic features from VGRF signals including swing time, stride time, and stride length can reveal the motor dysfunction and severity rating of PD. However, gait analysis using VGRF time series data pose several challenges including high dimensionality, nonlinear data dependency and complex correlation among kinematic features. Hence, to deal with the aforementioned challenges, quite a few a number of machine learning (ML) algorithms have already been investigated in previous work including support vector machine (SVM), linear discriminant analysis (LDA), decision tree (DT), principal component analysis (PCA) and random forest, (RF) etc. (Balaji et al., 2020; Khouri et al., 2019; Alam et al., 2017; Joshi et al., 2017). Nevertheless, the major limitation of ML techniques is the need for handcrafted feature vector. Therefore, recently deep learning (DL) algorithms have recently gained widespread attention in PD diagnosis for their substantial advancements and successful implementations in several fields including autonomous vehicles, cyber security, and natural language processing (Li et al., 2020).

In the literature, there have been considerable efforts to explore the efficacy of deep learning models for PD diagnosis. For instance, El Maachi et al. (2020) investigated the PD multi-class classification problem using CNN network and achieved an accuracy of 85.3% in PD severity rating. Balaji et al. (2021b) put forward a gait analysis based LSTM network to identify not only the presence of PD but also the severity rating. In another study, Zhao et al. (2018b) presented a two channel model by combining CNN and LSTM to extract the spatiotemporal features from gait data to implement multi-class classification of PD. Even though the aforementioned previous studies quantified the stages of PD using deep learning classifiers, they did not explore the optimal number of VGRF channels to extract the transformed gait features that could minimize the computational burden of the classifier models. Hence, this work investigates the optimal number of VGRF signals through the gait variability analysis and implements the adaptive data driven decomposition technique called EMD to extract the prominent IMFs based on spectral analysis to train CNN-LSTM classifier model (Mei et al., 2021; Zhao et al., 2019).

The primary reason for employing the hybrid CNN-LSTM classifier model is as follows. CNN is widely used for gait classification for its potential to extract informative and prominent biomarkers from VGRF signals. However, the gait cycle has strong temporal dependencies and CNN can only explore the spatial correlations. In this direction, LSTM can extract the discriminative temporal information from the gait time series data. Hence, capitalizing on the potentials of LSTM network that can capture the long term temporal dependencies from time series data, we put forward a novel data-driven PD prediction model using hybrid CNN-LSTM network by exploiting the spatiotemporal features in the gait cycle. To identify the optimal VGRF signals that can serve as efficient inputs to classify the healthy controls and PD subjects, we perform the fluctuation amplitude variability analysis and choose the optimum VGRF channels. Moreover, as the gait rhythm fluctuations manifest nonstationary and nonlinear characteristics, we employ the EMD technique, which is data driven, multi-scale and locally adaptive, to represent the gait features in frequency domain. By decomposing the VGRF signals using the EMD technique into IMFs, the spectral analysis is performed to extract the dominant IMF that capture most of the energy from gait cycle. The dominant IMFs obtained from the optimal VGRF signals are utilized to form the input matrix to train the CNN-LSTM network. Subsequently, CNN is employed to extract the significant features from the IMFs and then the extracted features are utilized in LSTM to predict the severity rating of PD. In this direction, the major contributions of this work can be summarized as follows.

- The VGRF signals are decomposed using the EMD technique and the significant IMFs that contain the prominent gait information are extracted using the power spectral analysis.
- The transformed VGRF signals in the form of dominant IMFs are utilized to train the CNN-LSTM hybrid model that can automatically extract both the spatial and temporal features for gait classification.
- To solve the stochastic objective functions in the classifier model, an Adam optimization algorithm is used. Moreover, to address the overfitting problem in CNN-LSTM model, the L2 regularization technique along with drop technique is employed. The PD stage classification accuracy of 98.32% is achieved using the proposed hybrid CNN-LSTM classifier model.

The rest of the article is structured as follows. Section 2 presents some of the related work on machine and deep learning methods for diagnosing and identifying the severity level of PD using gait signals. Section 3 gives the demographic details of the gait dataset considered for this study. Section 4 explains the proposed CNN-LSTM PD severity framework. Section 5 presents the results of proposed approach along with the performance analysis. Finally, Section 6 presents the conclusions.

2. Related work

Several diagnosis techniques have been put forward for not only identifying the presence of PD but also predicting the stages of PD based on the accepted unified Parkinson's disease rating scale (UPDRS) and Hoehn & Yahr (H&Y) scale. We present a survey of a few of the notable gait based PD diagnosis techniques proposed in the last decade in Table 1. This survey of existing techniques is divided into three classes of gait based PD prediction models: 1. statistical 2. machine learning and 3. deep learning based models. In the first category, the statistical parameters such as mean, median, standard deviation, kurtosis, energy and skewness are adapted as feature vector for classifying the healthy subjects and PD patients. However, the volume of data generated for gait analysis is large and the statistical methods require significant computational effort to diagnose the movement disorder. Moreover, the shortfalls of model driven prediction models including low accuracy with complex nonlinear data and limited robustness to data outlier have promoted significant interest in the data-driven prediction models. Hence, considerable research attention has also been paid to ML based approaches for PD diagnosis. Nevertheless, the need for handcrafted feature extraction in ML algorithms is a major limitation, which fueled further research on data-driven deep learning models for PD diagnosis. The DL models have the ability to automate the selection of feature vector and minimize the human intervention in solving the classification problem significantly (Zhao et al., 2019; Zhang et al., 2021). Considering the limitations of the previous techniques and research, we put forward a novel data-driven PD prediction model using hybrid CNN-LSTM network by exploiting the IMFs of selected VGRF signals through EMD technique and power spectral analysis.

3. Dataset

We utilize the gait time series dataset from the Physionet database (Gait in Parkinson's Disease, 2021) to evaluate the efficacy of the proposed scheme. The dataset, which was contributed by three independent research groups for various walking tests, was collected at the Tel-Aviv medical center, Israel. To acquire the foot pressure as a function of time, 8 foot sensitive resistors (FSR) were attached to each foot and the data were acquired while the subjects walked at their regular pace. Fig. 1 shows the positioning of the FSR along with their coordinates. The VGRF signals were collected for two minutes at a sampling rate of 100 Hz. The dataset consists of VGRF record of 93 persons affected with PD (PPD) and 73 healthy controls.

Table 1

Related works on gait based PD diagnosis and severity rating.

Category	Author	Year	Method	Remarks
Statistical	Wu and Krishnan (2009)	2010	SVM	<ul style="list-style-type: none"> Employed non-parametric Parzen-window technique to evaluate the gait variability in PD patients. Validated the performance of SVM using leave one out cross validation technique and achieved 90.32% accuracy.
	Lee and Lim (2012)	2012	FNN	<ul style="list-style-type: none"> Extracted wavelet based features from the gait cycle Trained FNN to classify idiopathic PD patients and achieved a maximum classification accuracy of 77.33%
	Perumal and Sankar (2016)	2016	LDA	<ul style="list-style-type: none"> Extracted spatiotemporal and kinetic features from gait cycle Studied both gait and tremor features and assessed their significance in early detection of PD
	Prabhu et al. (2018)	2018	PNN	<ul style="list-style-type: none"> Extracted the prominent features through Hill-climbing feature selection technique and statistical analysis Used gait signals of 13 subjects to classify neurodegenerative diseases including Huntington, PD, and Amyotrophic Lateral Sclerosis.
	Yurdakul et al. (2020)	2020	NR-LBP	<ul style="list-style-type: none"> Transformed raw VGRF signal using neighborhood representation local binary pattern (NR-LBP) for feature extraction Used Student's t-test to form different feature sets and obtained an accuracy of 98.3%
ML	Wahid et al. (2015)	2015	RF, SVM	<ul style="list-style-type: none"> Used multiple regression normalization strategy to find the differences in spatiotemporal gait parameters Achieved the highest classification accuracy of 92.6% using random forest technique
	Alam et al. (2017)	2017	SVM, KNN, RF, DT	<ul style="list-style-type: none"> Employed sequential forward feature selection Approach Identified that SVM with cubic kernel function yields best classification accuracy of 93.6% and sensitivity of 93.1%
	Joshi et al. (2017)	2017	SVM	<ul style="list-style-type: none"> Showed that wavelet combined with SVM can effectively differentiate between healthy control and PD subjects Identified that left stance and right swing produced maximum accuracy
	Khoury et al. (2019)	2019	KNN, DT, BC, SVM	<ul style="list-style-type: none"> Evaluated the gait pattern disorders using supervised ML algorithms to classify neurodegenerative diseases Utilized wrapper method based on RF method to identify the prominent features
	Balaji et al. (2020)	2020	SVM, EC, DT, BC	<ul style="list-style-type: none"> Performed statistical analysis of VGRF time series data to identify the significant features Compared the performance of four ML algorithms and achieved maximum classification accuracy of 99.4%
DL	Zhao et al. (2018b)	2018	CNN-LSTM	<ul style="list-style-type: none"> Designed a two channel model that using CNN combined with LSTM to learn the gait spatiotemporal patterns Achieved the highest classification accuracy of 98.80% using the hybrid model
	Ashour et al. (2020)	2020	LSTM	<ul style="list-style-type: none"> Adopted LSTM network to detect the freezing of gait (FOG) from acceleration Signals Extracted hybrid DWT-FFT features and trained SVM for performance comparison
	El Maachi et al. (2020)	2020	CNN	<ul style="list-style-type: none"> Used 1D CNN to design a DNN classifier for multi-class classification Employed ten-fold cross validation technique to avoid data overfitting and achieved 98.7% accuracy.
	Oktay and Kocer (2020)	2020	LSTM	<ul style="list-style-type: none"> Utilized a LSTM classifier to distinguish between Parkinsonian tremor (PT) and essential tremor (ET) Utilized CNN for feature extraction and obtained 90% accuracy for binary classification
	Balaji et al. (2021b)	2021	LSTM	<ul style="list-style-type: none"> Designed a LSTM classifier for severity rating of PD according to H&Y scale Employed Adam optimizer to solve the stochastic cost function and achieved average classification accuracy of 96.6%

Abbreviations used in the table are: FNN—Fuzzy neural network, PNN—Probabilistic neural network, LDA—Linear discriminant analysis, DT—Decision tree, RF—Random forest, BC—Bayes classifier, EC—Ensemble classifier, DWT—Discrete wavelet transform, FFT—Fast Fourier transform, DNN—Deep neural network.

Table 2 gives the demographics of the healthy control and PD subjects. The three datasets contributed by Yogeve et al. (2005), Hausdorff et al. (2007), and Frenkel-Toledo et al. (2005), consist of VGRF time series data for three walking tests namely 1. walking on a level ground

while performing the dual task 2. walking with rhythmic auditory stimulation and 3. Treadmill walking. **Table 3** presents the number of PD subjects participated in each walking test along with their severity rating. Since the dataset also consists of the severity level of PD based

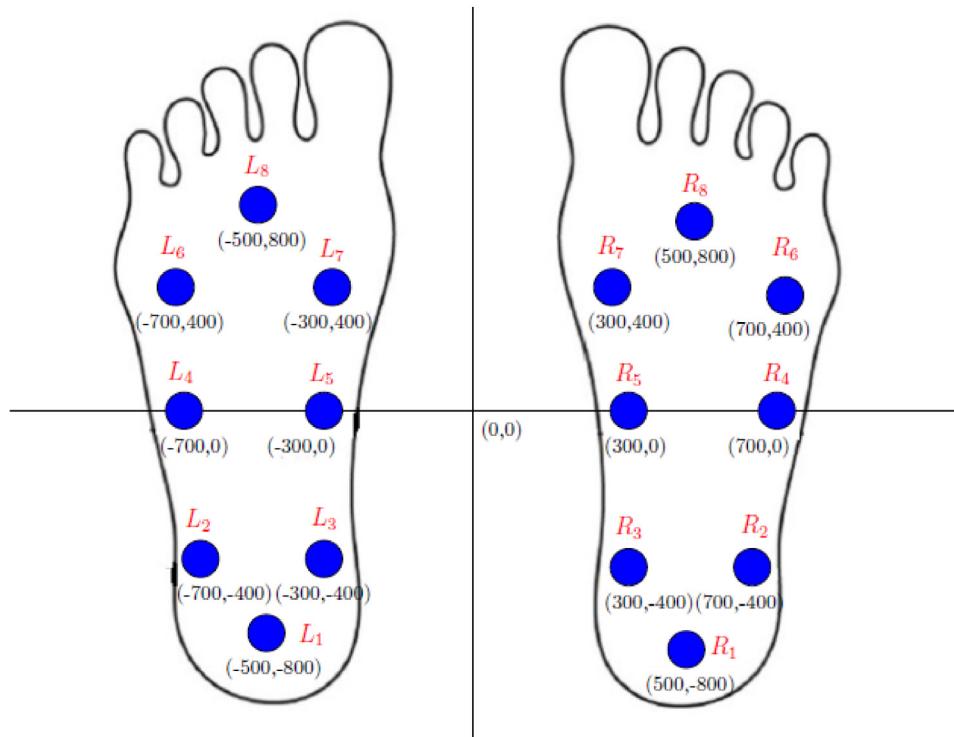
Fig. 1. FSR sensor positioning in $x - y$ coordinate system.

Table 2

Demographic details of subjects who volunteered for VGRF data collection.

Dataset	Category	No of subjects		Age (yr)			Height (m)	Weight (kg)
		Male	Female	<50	50~70	≥70		
Ga Frenkel-Toledo et al. (2005)	Healthy	10	8	0	10	8	1.68 ± 0.09	74.2 ± 12.6
	PD subject	20	9	0	13	16	1.67 ± 0.06	73.1 ± 11.3
Ju Hausdorff et al. (2007)	Healthy	12	14	0	20	6	$1.83 \pm .07$	66.8 ± 11.08
	PD subject	16	13	1	18	10	$1.87 \pm .16$	75.1 ± 16.90
Si Yogeve et al. (2005)	Healthy	18	11	1	26	2	$1.69 \pm .07$	71.5 ± 11.03
	PD subject	22	13	0	28	7	$1.66 \pm .08$	70.3 ± 8.35

Table 3

Ground truth of H&Y severity rating of PD subjects in three datasets.

Stage	Ga (Frenkel-Toledo et al., 2005)	Ju (Hausdorff et al., 2007)	Si (Yogeve et al., 2005)
Healthy	18	26	29
Severity 2	15	12	29
Severity 2.5	8	13	6
Severity 3	6	4	0

on H&Y scale, we can assess the efficacy of the proposed classifier scheme by considering the severity level given in the databank as the ground truth. Fig. 2 shows the exemplary cumulative VGRF signals from left and right feet of healthy and PD subjects. We can note that compared to healthy controls, in PD subjects the stride to stride variability is significantly high. Furthermore, the force magnitude of PD subjects while walking is relatively less than that of the healthy controls.

3.1. Spatiotemporal features

Fig. 3 shows the various phases of a gait cycle. The gait cycle can generally be categorized into the two major phases of swing and stance. The stance phase, which constitutes around 60% of the gait cycle, represents the time period in which foot maintains a firm contact with

the ground. Similarly, the swing phase which occupies the remaining 40% of the gait cycle is the time period the foot is off the ground (Balaji et al., 2021a). Further, the gait phase can be grouped into 6 events such as heel strike, mid-stance, foot flat, heel-off, mid-swing and toe-off. Some of the prominent spatial features are: stride length, step length, gait speed, and cadence. Likewise, the primary temporal features are: stride time, step time, stance time and swing time. In case of ML based approaches, by analyzing these phases, the significant spatial and temporal features of gait can be identified to assess the gait impairment. However, in this work, we employ the CNN for automatically extracting the significant gait features from the decomposed IMFs for PD classification. In the next section, we briefly explain the architecture of the deep learning framework and discuss the functionality of each module in the deep learning classifier model.

4. Methodology

Fig. 4 shows the proposed EMD based CNN-LSTM framework for PD severity prediction. The VGRF signals acquired using FSRs are integrated to form the gait dataset. Firstly, to minimize the influence of startup and endup effects in VGRF signals, the 10 s samples from the starting and the 20 s samples at the end are discarded. Then, to smoothen the gait signal and remove the outliers, the VGRF data are filtered using a median filter.

The filtered VGRF signals are analyzed for computing their variability and signal strength to identify the optimal number of VGRF

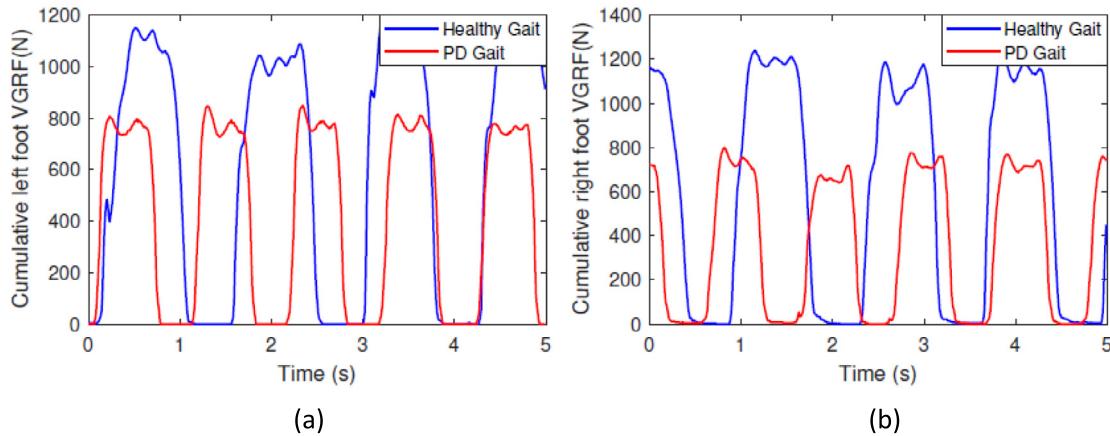


Fig. 2. Gait cycle of healthy and PD subjects (a) Left foot (b) Right foot.

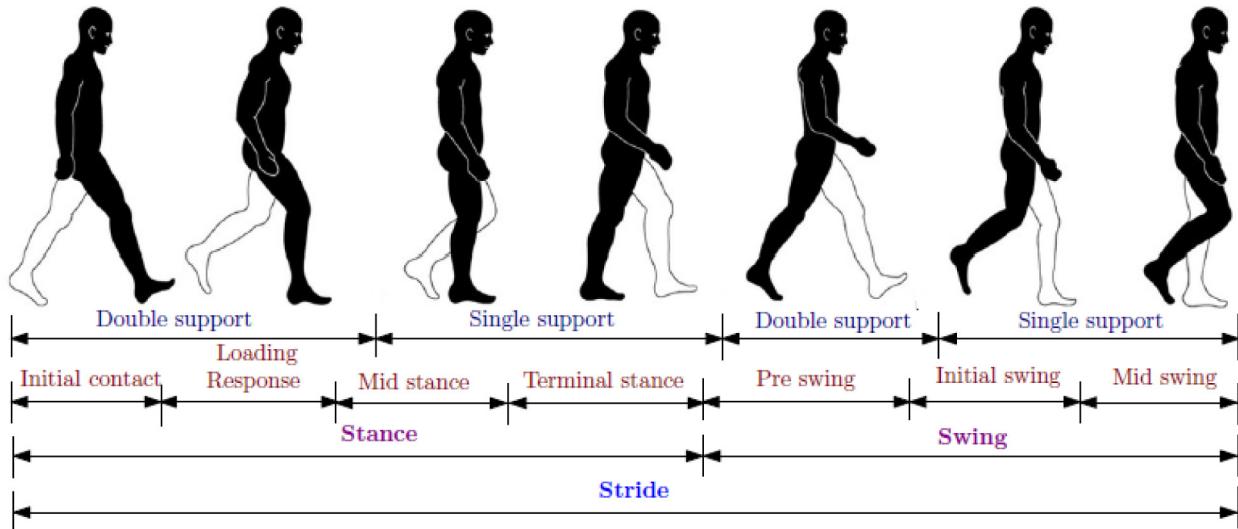


Fig. 3. Gait cycle.

channels that can be used to extract the dominant IMFs to train the classifier model. In the following section, we briefly discuss the various stages involved in identifying the stages of PD through gait variability analysis.

4.1. Optimal selection of VGRF signals

As the gait cycle duration of PD subjects is longer than that of the healthy controls, we propose to utilize only limited number of FSR sensors that can significantly reveal the stride-to-stride variability. Hence, we determine the fluctuation amplitude variability (FAV) of VGRF sensors in both the legs, as given in 1, and select the sensors that have the largest FAV and the highest force amplitude. Table 4, which shows the FAV of the sensors, manifests that sensors 3 and 7 have the largest variability. Fig. 5, which shows the sample VGRF signals of eight left foot sensors, highlights that the amplitudes of signals from sensors 1 and 8 are the largest. Hence, the VGRF signals from four left leg sensors (L1,L3,L7,L8) and the corresponding four right leg sensors (R1,R3,R7,R8) are considered for analyzing the gait dynamics and extracting the IMFs that can be used to train the deep learning models. As the gait signals are non-stationary, this work employs EMD technique to decompose the VGRF signals to not only denoise but also identify the modes of oscillations from IMFs.

$$FAV = \frac{|L_k - R_k|}{L_k} \times 100\% \quad (1)$$

Table 4
FAV of VGRF sensors.

Sensor	1	2	3	4	5	6	7	8
FAV	0.71	0.16	2.51	1.31	2.21	1.68	2.26	0.89

4.2. EMD technique

EMD, put forward by Norden E. Huang and his group, is an adaptive data-driven technique which is highly suitable for nonstationary and nonlinear data analysis. The key advantage of EMD is that through a sifting process it can decompose the signal into several IMFs and can capture the short time variations in frequencies that cannot normally be obtained using Fourier spectral analysis (Arivazhagan et al., 2021). The sifting operation assists in smoothing the uneven magnitudes in the signal of interest. Furthermore, compared to Wavelet and Fourier analysis, EMD technique does not require any prior knowledge of the basis of the system and works solely on data-driven decomposition approach (Vican et al., 2021). Through an iterative process, EMD technique obtains the IMFs of the nonlinear/nonstationary signal by repeatedly subtracting the average of lower and upper envelope from the original signal. As IMFs are subtracted from the original signal, the frequency of every new IMF will be lesser than the IMFs created before it. Consequently, this process provides an option to analyze the different modes of oscillation in the signal of interest and helps to

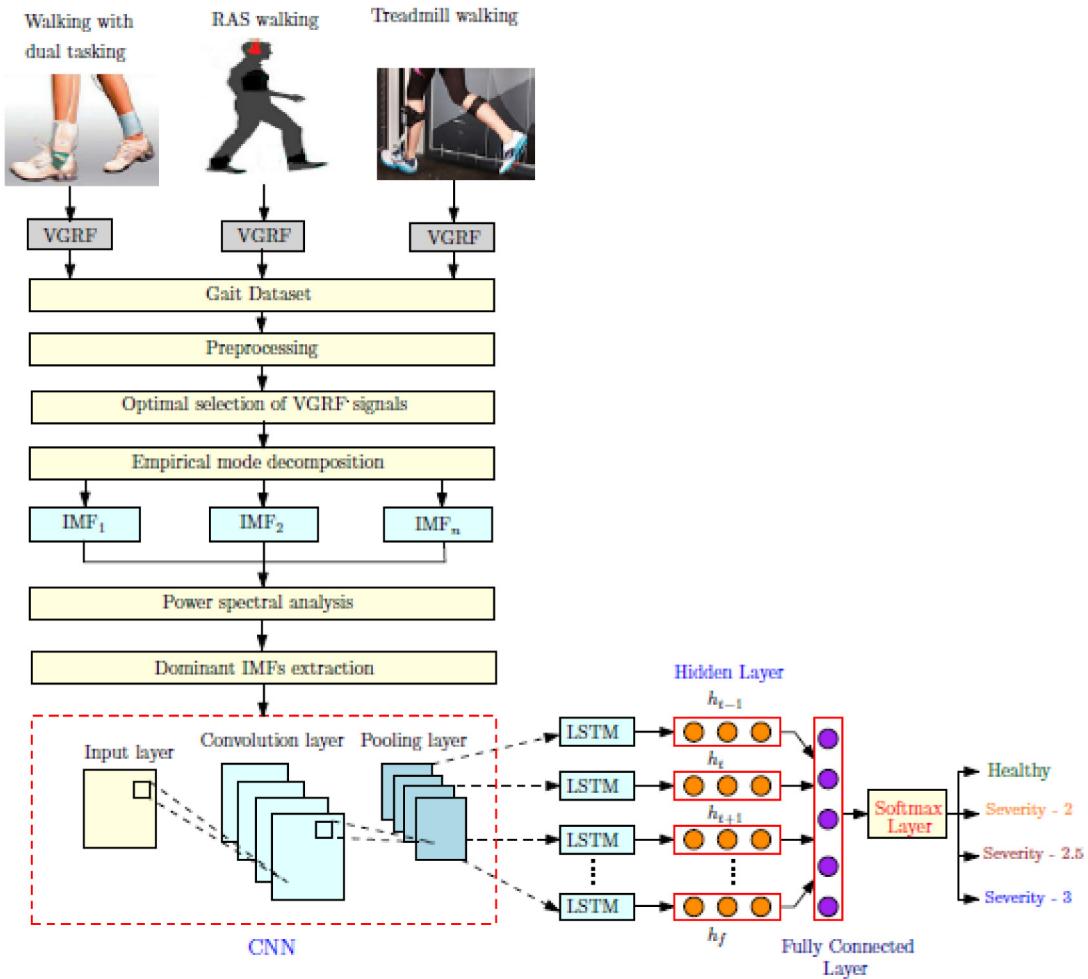


Fig. 4. Proposed CNN-LSTM architecture for PD severity rating using gait cycle.

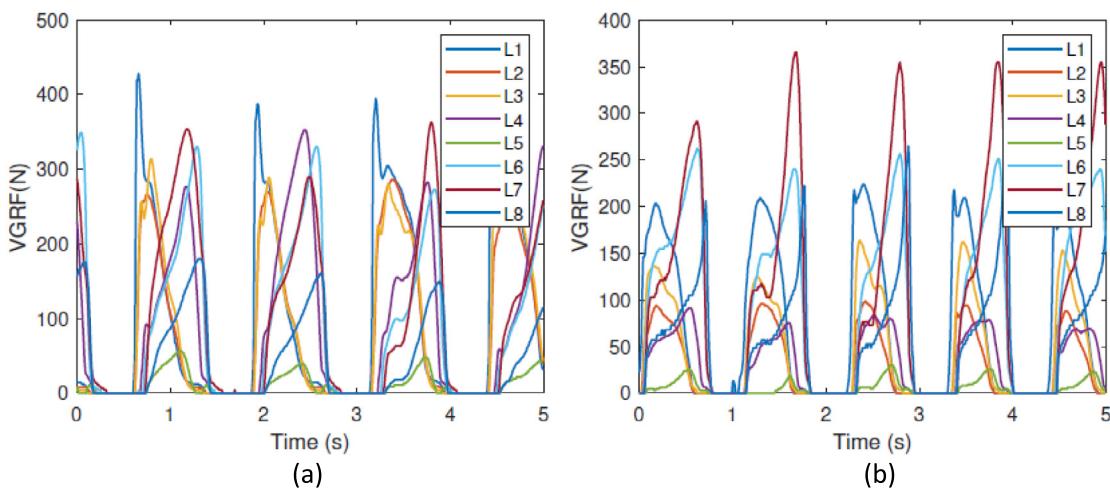


Fig. 5. Exemplary VGRF signals of left foot sensors of (a) Healthy (b) PD subjects.

extract the prominent features from IMFs. Every IMF generated from the sifting process should satisfy the following two conditions to be saved as an IMF. i. The difference between the number of extrema and zero crossings should be either 0 or 1. ii. The average of lower envelope and upper envelope must be 0. Typically, to avoid losing important amplitude information due to large of number of iterations in EMD sifting process, stopping criteria that can limit the number of

iterations are used. For instance, until the residue $r(t)$ turns out to be a monotonic function from which no further IMF could be obtained. Algorithm 1 presents the pseudo code of the EMD technique.

The power spectral obtained from the IMFs of the selected VGRF signals are determined to find the dominant IMFs that contain the significant gait information. Then, the respective dominant IMFs are used as inputs to train the CNN-LSTM classifier model.

Algorithm 1 EMD algorithm

1: Input : VGRF time series data ($x(t)$)
2: Output : A group of IMFs ($imf_i(t)$) and a residue ($r_i(t)$)
3: Initialize : Set IMF index with $i = 1$
4: Find the local minima and maxima points in the signal $x(t)$.
5: Through cubic spline interpolation, find the lower envelope $e_l(t)$ and the upper envelope $e_u(t)$.
6: Compute the average envelope $e_a(t)$

$$e_a(t) = \frac{e_l(t) + e_u(t)}{2}$$

7: Find the detail signal by subtracting average envelope $e_a(t)$ from $x(t)$

$$d(t) = x(t) - e_a(t)$$

8: If $d(t)$ meets the two IMF conditions, then
9: Assign i^{th} IMF as $imf_i(t) = d(t)$
10: Else
11: Set $x(t) = d(t)$ and move to step 4.
12: End if
13: Assign i^{th} residue $r_i(t) = x(t) - imf_i(t)$
14: Update $x(t) = r_i(t)$
15: $i = i + 1$
16: If $r(t)$ is a not a monotonic or a constant
17: Go to step 4 and repeat
18: Else
19: Stop the computation and return all IMFs and residue.

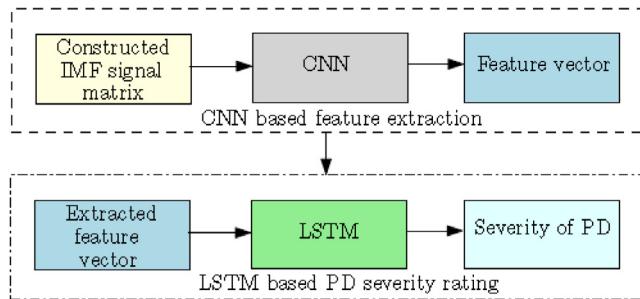


Fig. 6. CNN-LSTM hybrid network.

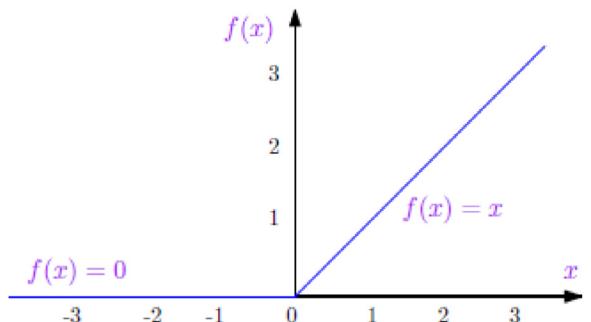


Fig. 7. ReLU activation function.

4.3. Feature extraction through CNN

The CNN-LSTM framework for PD severity rating consists of cascade connection of CNN and LSTM. The CNN, as illustrated in Fig. 6, can extract the deep features from the input IMFs matrix and LSTM can classify the gait pattern based on the deep features extracted from CNN. The top layer of CNN-LSTM consists of CNN layer that receives the IMFs that contain the significant gait information from three walking tests.

Typically, a CNN contains three layers, namely the input, hidden and output layers (Sharma et al., 2021; Kim and Cho, 2019). The hidden layer generally contains a convolution layer and a pooling layer along with a ReLU layer which constitutes an activation function. The convolutional layer moves a set of learnable filters over the VGRF input based on the kernel function to extract the prominent features. Convolutional operation can minimize the number of parameters and enable the hybrid network deeper. If $x_i^0 = x_1; x_2; \dots; x_n$ is the input

vector, the feature mapping process can be represented as

$$y_{ij} = f \left(\sum_{m=1}^M w_{m,j} x_{i+m-1,j}^0 + b_j \right) \quad (2)$$

where M, w, b indicate the kernel size, weight of the feature map and bias term, respectively. f is the activation function, which is considered to be ReLU in this application mainly to introduce the nonlinearity in the layer as follows.

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \quad (3)$$

The ReLU, as shown in Fig. 7, is a piecewise linear function that provides 0 as an output if it receives negative input value and the value itself if it gets positive input value. The important feature of the ReLU is that the rate of convergence is faster than the other nonlinear activation functions such as σ and $tanh$.

The pooling layer, also called subsampling layer, implements a downsampling operation to minimize the feature size and assures translation invariance (Chen et al., 2020). Through small rectangular blocks,

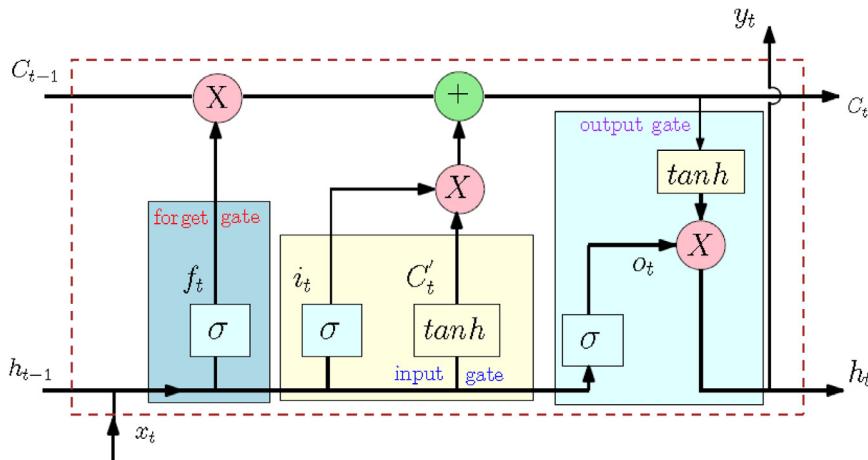


Fig. 8. LSTM architecture.

it results in singular output for each block. Max and average pooling are the two widely used pooling operations. In this study, we utilize the following Max pooling operation to downsample the feature map.

$$p_i^{l,j} = \max_{r \in R} c_{iX T+r}^{l,j} \quad (4)$$

where R represents the pooling window size and T indicates the pooling stride.

4.3.1. LSTM for classification

LSTM network, put forward by Hochreiter and Schmidhuber (1997), is a class of recurrent neural network (RNN) that can extract the long-term dependencies from the temporal sequences. The two major advantages of LSTM network over conventional RNN are as follows: (a) LSTM can solve the gradient vanishing and exploding problems in conventional RNNs using the memory blocks (b) For their sequential structure, LSTM is powerful to obtain the temporal attributes from the time series data (Zhao et al., 2019). The LSTM architecture, as shown in Fig. 8, contains four gates, namely forget gate (f_t), input gate (i_t), update gate (o_t) and output gate (o_t) along with a memory unit called a cell (Hussein et al., 2019).

The update of LSTM units at every timestep t is governed by the following equations.

$$f_t = \sigma(W_{hf}h_{t-1} + W_{xf}x_t + b_f) \quad (5)$$

$$i_t = \sigma(W_{hi}h_{t-1} + W_{xi}x_t + b_i) \quad (6)$$

$$o_t = \sigma(W_{ho}h_{t-1} + W_{xo}x_t + b_o) \quad (7)$$

$$g_t = \tanh(W_{hc}h_{t-1} + W_{xc}x_t + b_c) \quad (8)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (9)$$

$$o_t = h_t \odot \tanh(c_t) \quad (10)$$

where W and b respectively indicate the weight matrices and bias vector. σ depicts the sigmoid function which can be represented as

$$\sigma(x) = (1 + e^{-x})^{-1} \quad (11)$$

4.3.2. Softmax layer

Softmax layer performs the prediction of class label using the extracted features. It represents the generalization of logistic regression to multi-class classification problem and can be represented as follows.

$$z_i = \sum_j h_j w_{ji} \quad (12)$$

$$\text{Softmax}(z_i) = p_i = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (13)$$

where z_i , h_j and w_{ji} indicate the input of softmax layer, activation function in the penultimate layer, and weight between the softmax and

penultimate layers, respectively. The predicted class label \hat{y} from the softmax layer will be as follows.

$$\hat{y} = \arg \max_i p_i \quad (14)$$

Algorithm 2 presents the pseudo-code of the CNN-LSTM algorithm for predicting the severity of PD. The dominant IMFs obtained based on the power spectral are given as the input to the CNNLSTM classifier model and the four classes of PD namely healthy (H), severity-2 (S-2), severity-2.5 (S-2.5) and severity-3 (S-3) are predicted using LSTM network based on the probabilistic information obtained from softmax layer.

4.4. Minimizing overfitting

4.4.1. L2 regularization

Regularization can improve the prediction performance by alleviating the overfitting problem. Typically, regularization shrinks the data coefficients close to zero so as to make the classifier model more flexible to learn even in a complex environment. In this work, we employ L2 regularization technique, also called ridge regression, to address the data overfitting. Typically, in L2 regularization, a squared magnitude of coefficient is added as a penalty factor to the loss function. Eq. (15) gives the loss function added with the L2 regularization term.

$$\hat{H}(\omega, Y) = H(\omega, Y) + \lambda R(\omega) = H(\omega, Y) + \lambda \sum_i |\omega_i|^2 \quad (15)$$

where $H(\omega, Y)$ indicates a cross-entropy loss function. λ represents the regularization coefficient which can influence the resultant loss function $\hat{H}(\omega, Y)$. Tuning the hyperparameter, called regularization coefficient, penalizes the parameters of the model. We can note that the resultant cost function will be convex because $R(\omega)$ is convex. Hence, penalizing the sum of squared weights through λ , L2 regularization makes the classifier model to yield minimum generalization error.

4.4.2. Dropout

Dropout is another regularization technique which randomly subsamples the output of a layer. Unlike L1 and L2 regularization methods that modify the cost function, the dropout alters the network by dropping units randomly during network training. The motivation for a dropout layer is that a fully connected layer may result in co-adaptation due to identical connection weights of two different neurons, resulting in both neurons extracting the same features from input data.

Algorithm 2 PD severity prediction using CNN-LSTM

```

1: Input : IMFs of VGRF time series data, trained CNN-LSTM model
2: Output : Predicted PD severity rating  $\hat{y} \leftarrow \{1, \dots, C\}$ 
3: Initialization : No of classes ( $C=4$ ,  $d=12000$ ,  $M=500$ )
4: procedure (PD-CNN-LSTM( $x, C, d, M$ , CNN-LSTM))
5: Segment the gait series and select the length ( $L$ )
6: Apply EMD technique and extract the IMFs
7: Perform the power spectral analysis of IMFs and select the dominant IMFs
8: while  $i \leq M$  do
9: Identify the feature map from the dominant IMFs
10: Perform pooling and downsample the feature map
11: Flatten the feature map for LSTM network
12:  $z_t = g(W_z x_t + R_z y_{t-1} + b_z)$   $\triangleright$  LSTM Input
13:  $i_t = \sigma(W_{hi} h_{t-1} + W_{xi} x_t + b_i)$   $\triangleright$  Input gate
14:  $f_t = \sigma(W_{hf} h_{t-1} + W_{xf} x_t + b_f)$   $\triangleright$  Forget gate
15:  $c_t = f_t \odot c_{t-1} + i_t \odot g_t$   $\triangleright$  Cell
16:  $o_t = \sigma(W_{ho} h_{t-1} + W_{xo} x_t + b_o)$   $\triangleright$  Output gate
17:  $o_t = h_t \odot \tanh(c_t)$   $\triangleright$  LSTM Output
18:  $s_t = h_t(y_t)$   $\triangleright$  FC Layer
19:  $E = AP(s_t, s_{t-1}, \dots, s_{t-M})$ ;  $\triangleright$  Average pooling
20: Compute  $P_C = \{P_1, \dots, P_c\} \leftarrow \text{Softmax}(E)$ 
21: Find IHP  $\leftarrow \text{Support}(\max(P_C))$   $\triangleright$  Index of highest probability
22:  $\hat{y} = \text{IHP}$ ;  $\triangleright$  Predicted PD severity rating

```

Hence, the duplicate extracted features that are particular to only the training set will lead to overfitting. To address this issue, in dropout, each neuron is dropped out with a probability of $1 - p$, resulting in reduced subnet. Then, the training is performed on the resultant subnet, thereby reducing the computational overhead. Eq. (16) gives the forward propagation augmented with dropout.

$$z_j = \sum_i W_{ij} d_i x_i + b_i \quad (16)$$

where d_i represents a vector which contains independent Bernoulli random variables with p probability. The input node will be discarded when d_i is zero.

4.5. Adam optimizer

For optimizing the stochastic objective functions, an adaptive moment estimation, also called Adam, algorithm is utilized because it needs minimal tuning and does not require memory. Moreover, Adam combines the advantages of two stochastic gradient descent algorithms namely Adaptive gradient and root mean square propagation algorithms and works well for sparse gradients issues. Utilizing the estimated moments to optimize the stochastic cost function, Adam effectively solves the non-convex optimization problem and is comparatively easy to customize. From an algorithm standpoint, it computes the exponential weighted moving average of the gradient and finds the square of the gradient. The two decay parameters β_1 and β_2 regulate the rate at which these moving averages are computed. Adam determines the decaying mean of past gradient μ_t and its squared gradient γ_t as follows.

$$\mu_t = \beta_1 \mu_{t-1} + (1 - \beta_1) g_t \quad (17)$$

$$\gamma_t = \beta_2 \gamma_{t-1} + (1 - \beta_2) g_t^2 \quad (18)$$

During the initial time steps, since μ_t and γ_t are set as null vectors, the moment estimates are biased about zero particularly if the decay rates are less. Subsequently, these biases are corrected through the first and second moment estimates as follows.

$$\hat{\mu}_t = \mu_t / (1 - \beta_1^t) \quad (19)$$

$$\hat{\gamma}_t = \gamma_t / (1 - \beta_2^t) \quad (20)$$

Then, based on the following update rule, the estimated moments are utilized to renew the parameters.

$$\theta_t = \theta_{t-1} - \alpha \cdot \hat{\mu}_t / (\sqrt{\hat{\gamma}_t} + \epsilon) \quad (21)$$

It is proven in the literature that selecting the exponential decay rates β_1 and β_2 around 1 with minimal tolerance parameter ϵ of about 10^{-7} is highly suitable for sparse gradient problems. Table 3 presents pseudo code of the Adam optimization algorithm.

5. Experimental results and discussion

The experiments are conducted using the Tensorflow with Keras library to implement the proposed CNN-LSTM classifier model. The time series dataset is divided into two parts with 80% data for training and 20% data for testing. The number of samples in three datasets Ga, Ju and Si is 13,000, 11,000 and 7500 respectively. Fig. 9, which illustrates the exemplary left sensor-1 data from three walking tests for both PD subjects and healthy controls, clearly manifests that as the level of PD severity increases, the magnitude of VGRF significantly decreases. For instance, the force magnitude of PD subjects with severity rating 3 is almost one fourth that of the healthy subjects. This clearly highlights that the higher severity level of PD not only affects the gait rhythm but also the stride-to-stride variability.

5.1. IMFs extraction

As the gait signal is non-stationary and nonlinear, we firstly apply the EMD technique and compute the modes of oscillations through the IMFs. Dividing the VGRF signals into segments of 500 data points, we apply EMD technique and decompose each segment into IMFs. Fig. 10 shows the sample decomposed IMFs and the residue obtained from the VGRF signals of healthy and PD subjects. It is worth noting that IMF1 and IMF5 have the highest and the lowest frequencies, respectively. As the choice of suitable number of IMFs play a vital role in signal analysis, in this study, the first four IMFs of every VGRF segment that have the significant frequency content are considered to extract the distinctive features. As discussed in Section 4.1, we select the prominent VGRF signals from the 8 FSR sensors (L1,L3,L7,L8,R1,R3,R7,R8) and extract IMFs of those signals through EMD. Subsequently, through power spectral analysis the IMF that has most of the energy concentrated is

Algorithm 3 Adam optimization

```

1: Initialize: step size hyperparameter  $\alpha$ , exponential decay rates  $\beta_1, \beta_2 \in [0,1]$ , stochastic loss function  $f(\theta)$ , tolerance factor  $\varepsilon$ .
2:  $\mu_0, \gamma_0, t \leftarrow [0,0,0]$                                      ▷ Initialize 1st and 2nd moment estimates
3: while until  $\theta_t$  has not converged do
4:    $t = t + 1$                                               ▷ Update timestep
5:    $g_t \leftarrow \Delta_\theta f_t(\theta_{t-1})$                          ▷ Determine gradient
6:    $\mu_t \leftarrow \beta_1 \mu_{t-1} + (1 - \beta_1) g_t$           ▷ Compute 1st moment estimate
7:    $\gamma_t \leftarrow \beta_2 \gamma_{t-1} + (1 - \beta_2) g_t^2$        ▷ Compute 2nd moment estimate
8:    $\hat{\mu}_t \leftarrow \mu_t / (1 - \beta_1^t)$                   ▷ Find unbiased estimate of 1st moment
9:    $\hat{\gamma}_t \leftarrow \gamma_t / (1 - \beta_2^t)$                   ▷ Find unbiased estimate of 2nd moment
10:   $\theta_t = \theta_{t-1} - \alpha \cdot \hat{\mu}_t / (\sqrt{\hat{\gamma}_t} + \varepsilon)$  ▷ Determine objective parameters
11:  Return  $\theta_t$                                          ▷ Return final parameters

```

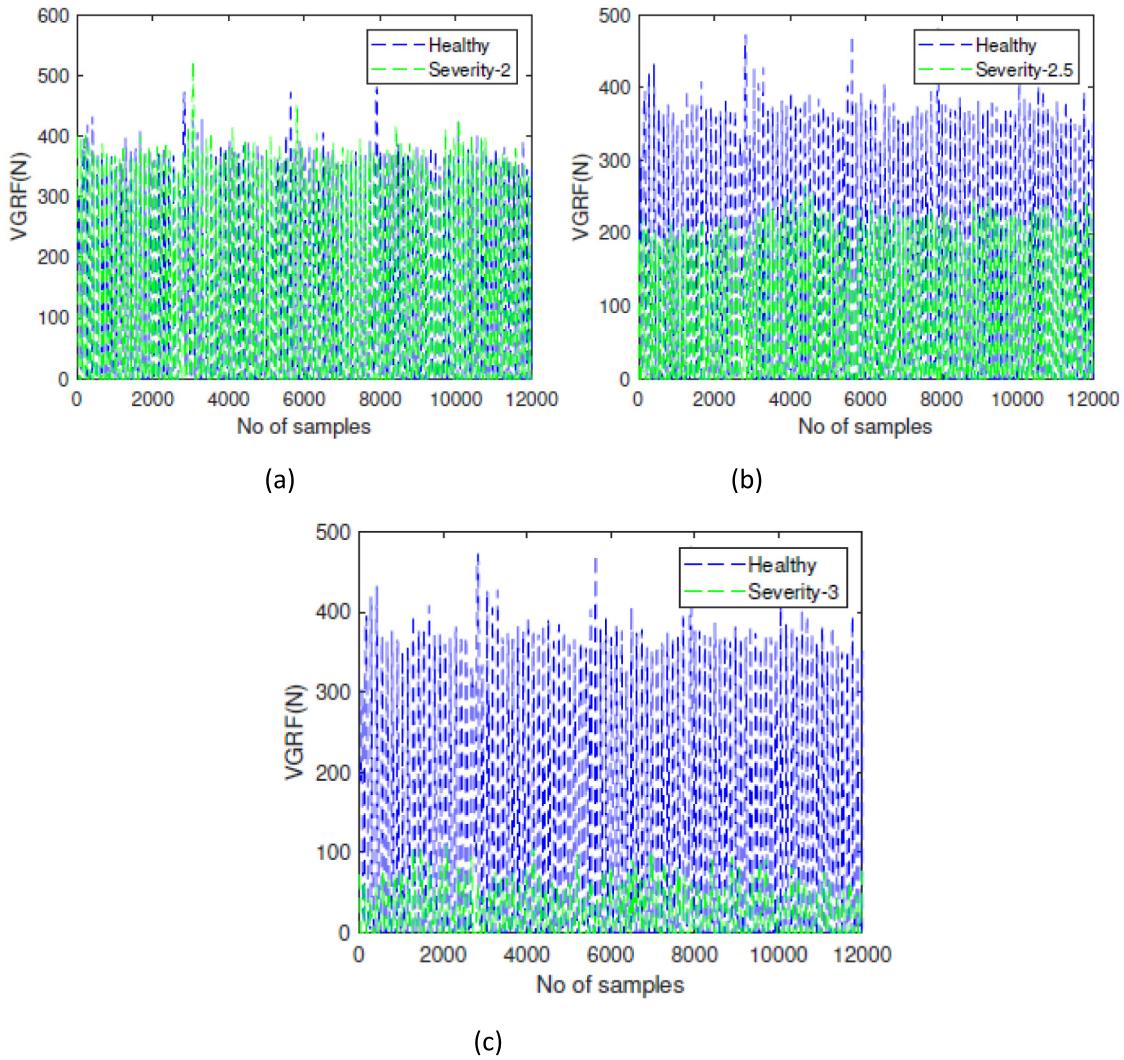


Fig. 9. SL1 signal for healthy and PD patients.

derived to reduce the computational burden of classifier model. Figs. 11 and 12 illustrate the sample power spectral plot of the first four IMFs extracted from the VGRF signals from L3 and L7 sensors, respectively for the four output classes namely healthy, S2, S2.5 and S3. We can note that second IMF has most of the energy contained across the class. Hence, the second IMF components from each of the eight VGRF signals are used to form the input matrix to train the deep learning model.

Table 5 gives the layer configuration of the CNN-LSTM network that consists of 4 convolutional and max-pooling layers alternatively

to extract the prominent features. The extracted features are provided as an input feature vector to LSTM network for the multi-class severity rating of PD.

First, the input layer is convolved with a kernel size of 250 to create layer1. Subsequently, the max pooling with a window size of 2 is applied onto each feature map. After 4 layers of sub-CNN architecture, 2 layers of LSTM and dropout layers are introduced with the former layer output being 50. Then, a fully connected layer followed by a softmax layer with layer output of 4, representing the four labels of PD

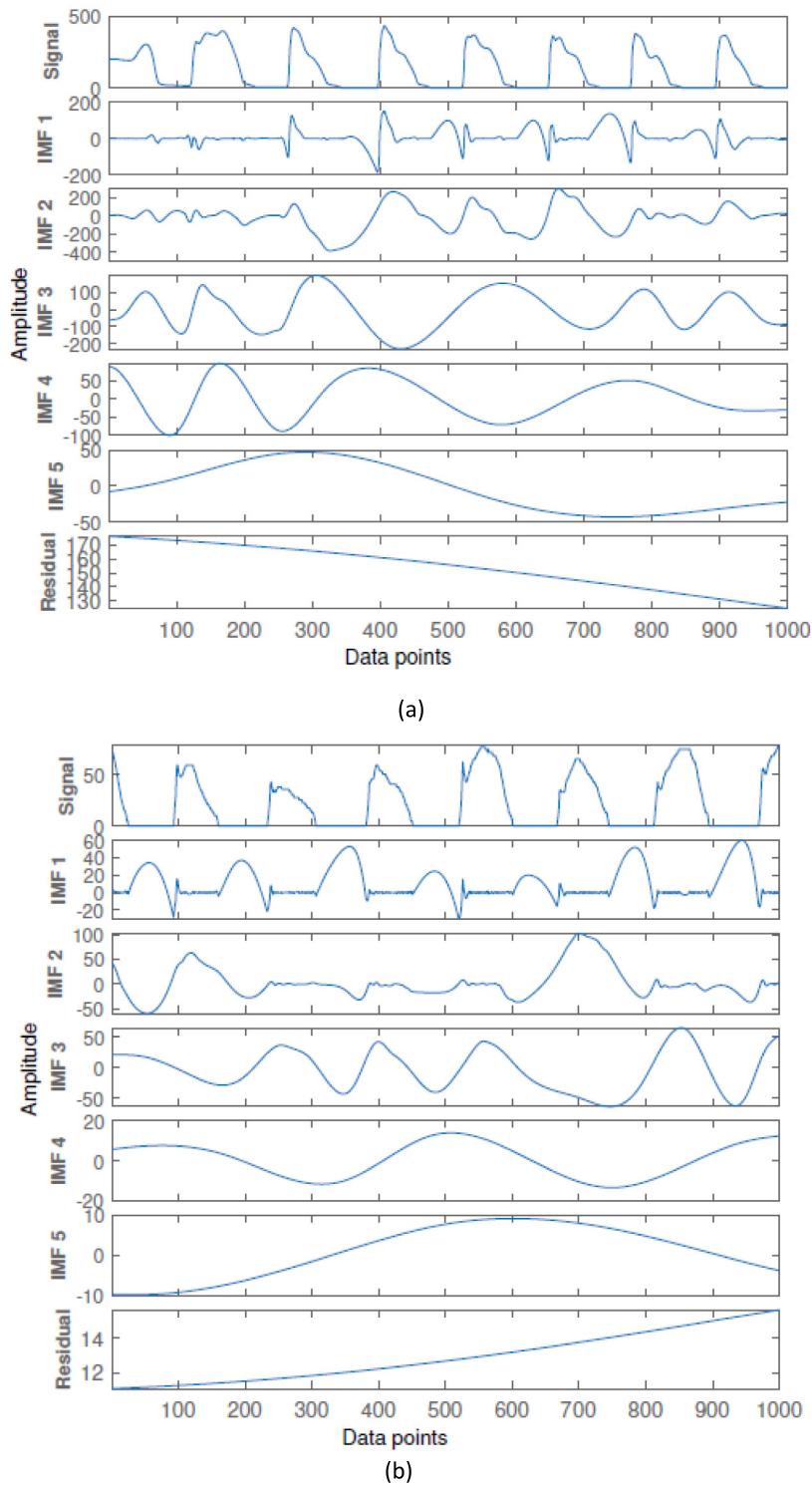


Fig. 10. EMD of VGRF signal (a) Healthy (b) PD subject.

classes in designed. The feature map flattened by the fully connected layer is inputted to the softmax layer to predict the class labels based on the probability value.

The DNN training is performed for 25 epochs, and the Adam optimizer is configured with the following hyper-parameter settings to minimize the loss function: the learning rate $\alpha = 0.001$, the exponential decay rates $\beta_1 = 0.9$, $\beta_2 = 0.99$, the threshold $\delta = 10^{-5}$ and constant $\epsilon = 10^{-8}$. While solving the cost function using Adam optimizer, to balance between the training error and learning speed of network, an adaptive learning rate approach is implemented. Hence, the Adam

optimizer starts with a learning rate of 0.001 and updates the weight iteratively until optimal results are obtained. Specifically, utilizing the 1st and 2nd moments of gradient, Adam updates the learning rate (α) for every weight.

5.2. Performance metrics

To evaluate the efficacy of the deep learning classifier model, the following performance metrics are used: Accuracy (Acc), specificity

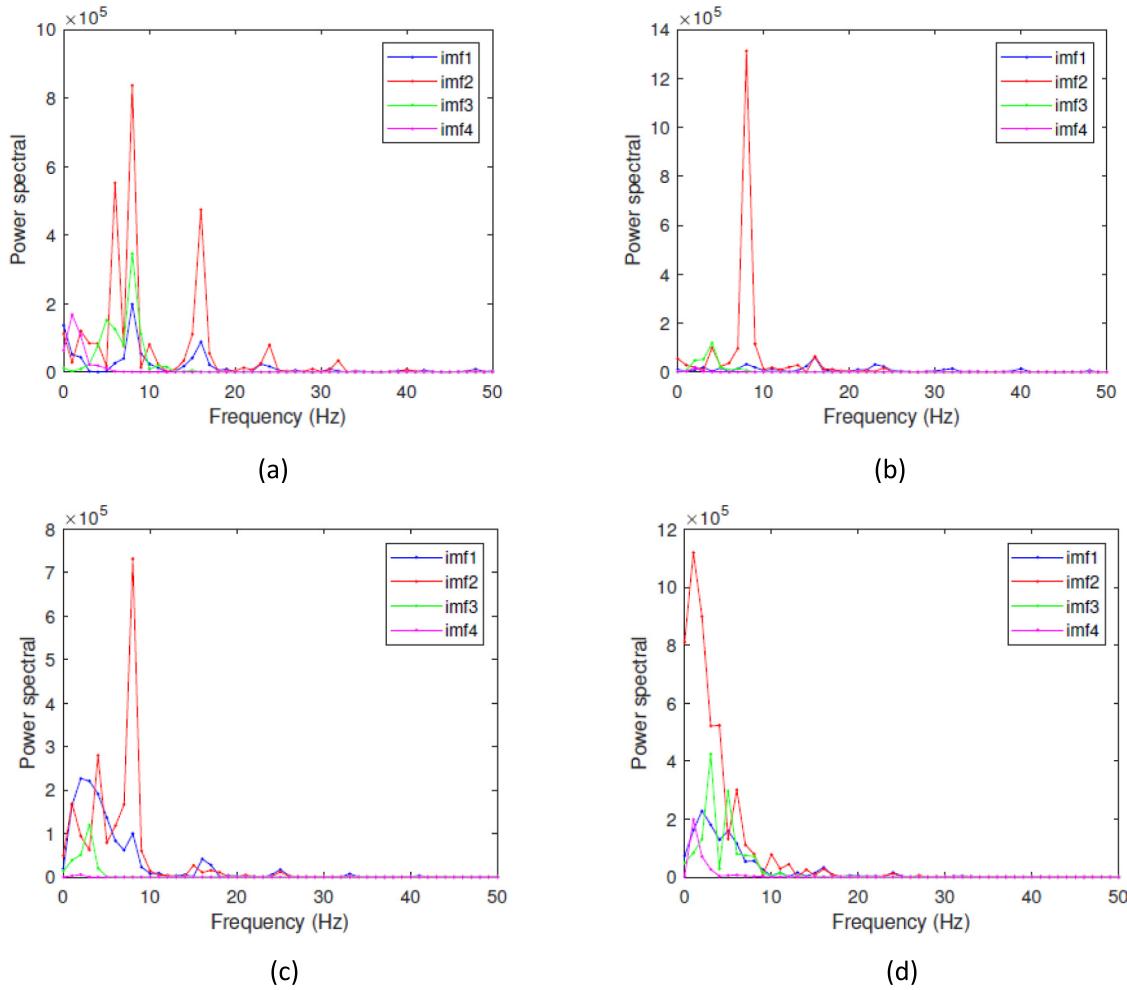


Fig. 11. Sample power spectral obtained from first four IMFs of VGRF at L3 (a) H (b) S2 (c) S2.5 (d) S3.

Table 5
CNN-LSTM layer configuration.

Layers	Filter	Activation	Layer output	Parameters	Stride
Convolution	12	ReLU	900 × 12	48	(1,1)
Pooling	2	–	450 × 12	0	(2,2)
Convolution	24	ReLU	450 × 24	888	(1,1)
Pooling	2	–	225 × 24	0	(2,2)
Convolution	48	ReLU	225 × 48	3504	(1,1)
Pooling	2	–	112 × 48	0	(2,2)
Convolution	96	ReLU	112 × 96	13920	(1,1)
Pooling	2	–	56 × 96	0	(2,2)
LSTM	–	–	50	10 400	–
Dropout (0.2)	–	–	–	–	–
LSTM	–	–	50	20 200	–
Dropout (0.2)	–	–	–	–	–
Fully connected	100	ReLU	50	86 450	–
Dense (25)	–	–	–	375 025	–
Dense (4)	–	Softmax	4	36	–
Total No. of parameters	–	–	–	446 129	–

(Spe), sensitivity (Sen), positive predictive value (PPV), F-score, and Mathew's correlation coefficient (MCC). The expressions to determine the metrics are as follows.

$$Acc (\%) = \frac{TN + TP}{TN + TP + FN + FP} * 100\% \quad (22)$$

$$Sen (\%) = \frac{TP}{TP + FN} * 100\% \quad (23)$$

$$Spe (\%) = \frac{TN}{TN + FP} * 100\% \quad (24)$$

Table 6
CNN-LSTM classifier performance metrics.

Parameters	CNN				CNN-LSTM			
	H	S-2	S-2.5	S-3	H	S-2	S-2.5	S-3
Acc (%)	97.42	97.70	98.42	97.20	98.84	98.76	99.29	99.18
Sen (%)	94.33	93.89	98.22	99.24	98.76	97.58	97.86	97.98
Spe (%)	98.46	99.02	98.48	99.01	98.87	99.16	99.32	99.58
PPV (%)	95.41	97.12	95.41	97.56	96.69	97.48	99.32	98.74
F-score (%)	94.86	95.48	96.7	98.39	97.71	97.50	98.58	98.35
MCC	0.93	0.94	0.94	0.96	0.97	0.96	0.98	0.96

$$F-score (\%) = \frac{Precision * Sensitivity}{Precision + Sensitivity} * 100\% \quad (25)$$

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (26)$$

5.3. Performance analysis

We have tested the proposed framework against the baseline framework with CNN to highlight the improvements in the classification performance. To examine the efficacy of the classifier, the confusion matrix, which provides the performance in the form of table layout, is shown in Fig. 13. From the confusion matrix, the six performance metrics such as accuracy, sensitivity, specificity, PPV, F-score and MCC are determined for both the classifier models. Table 6 gives the cumulative performance of the CNN and CNN-LSTM classifier models.

The CNN classifier has achieved the highest classification accuracy of 98.42% for the class severity-2.5. The misclassification rate in the

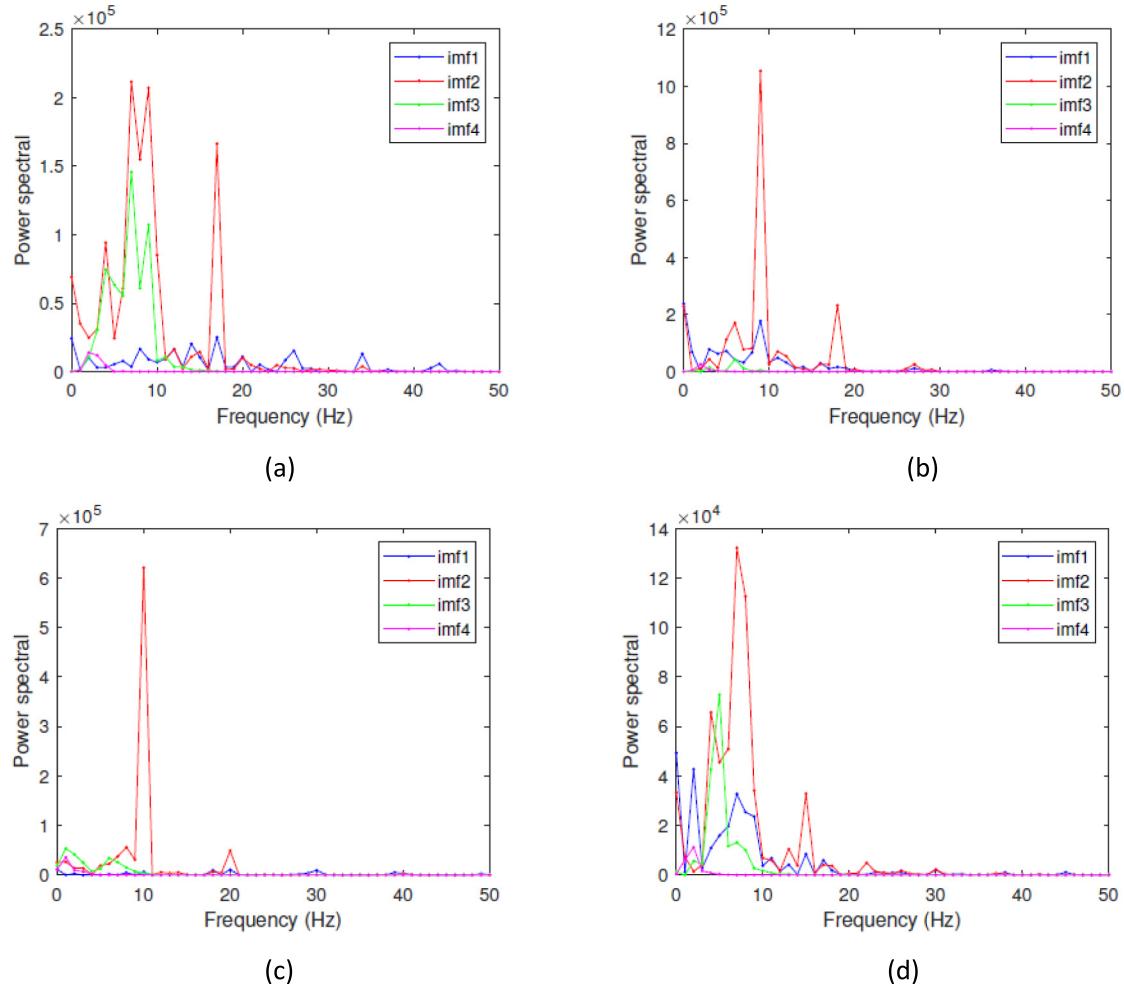


Fig. 12. Sample power spectral obtained from first four IMFs of VGRF at L7 (a) H (b) S2 (c) S2.5 (d) S3.

		Predicted label			
True label		H	S-2	S-2.5	S-3
		H	95.41	2.06	3.67
S-2	2.95	97.12	0.92	2.44	
S-2.5	1.31	0.41	95.41		
S-3	0.33	0.41		97.56	

True label		H	S-2	S-2.5	S-3
		H	98.76	1.21	
S-2	1.04	97.58	0.16	1.22	
S-2.5	0.83	1.31	97.86		
S-3	1.51		0.51	97.98	

(a)

(b)

Fig. 13. Confusion matrix (a) CNN (b) CNN-LSTM..

case of healthy subjects for CNN is 2.58%. However, the hybrid classifier has attained the highest accuracy of 99.29% for the same class (severity-2.5). Moreover, compared to CNN classifier, the misclassification across the four classes is very minimal in the CNN-LSTM model and the reason for the enhanced performance could be attributed to the

potential of LSTM to extract the temporal information from the input sequence.

Figs. 14 and 15 illustrate the performance of the classifier models during training and validation through the accuracy plot and loss function plot, respectively. The accuracy plot highlights that the L2

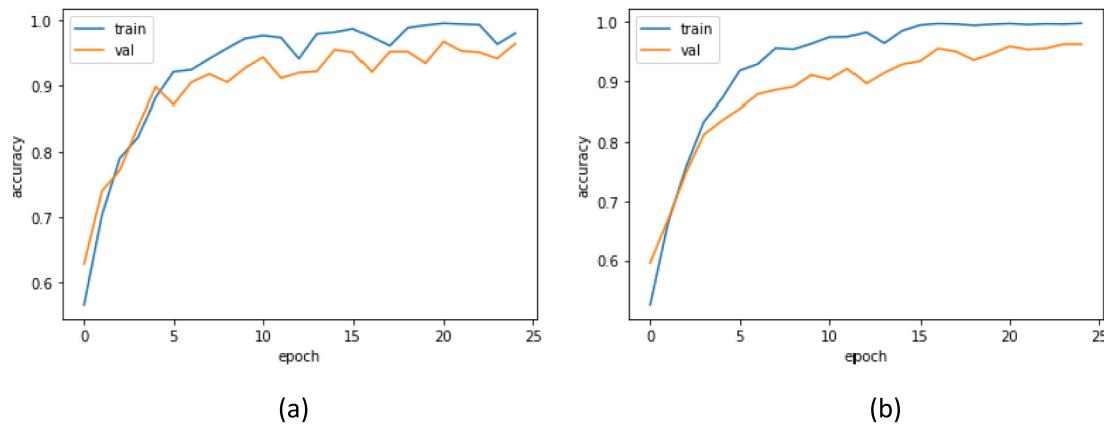


Fig. 14. Accuracy plot of (a) CNN (b) CNN-LSTM..

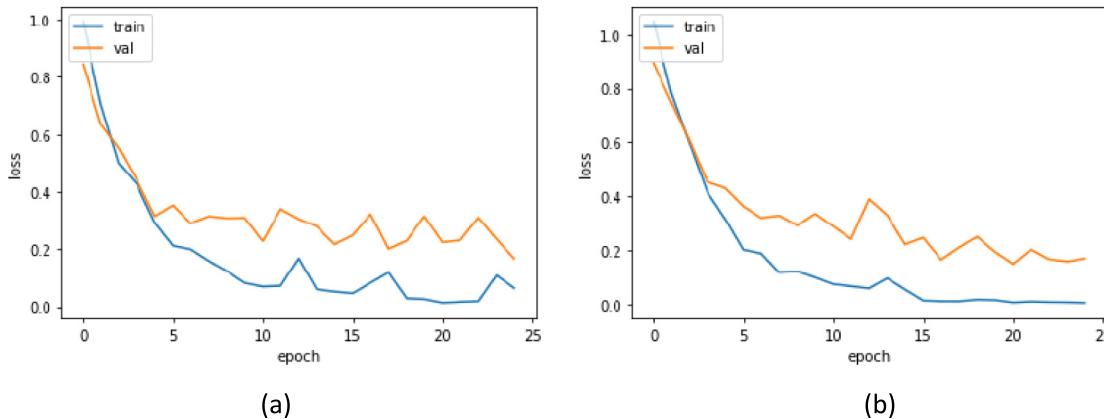


Fig. 15. Loss function plot of (a) CNN (b) CNN-LSTM..

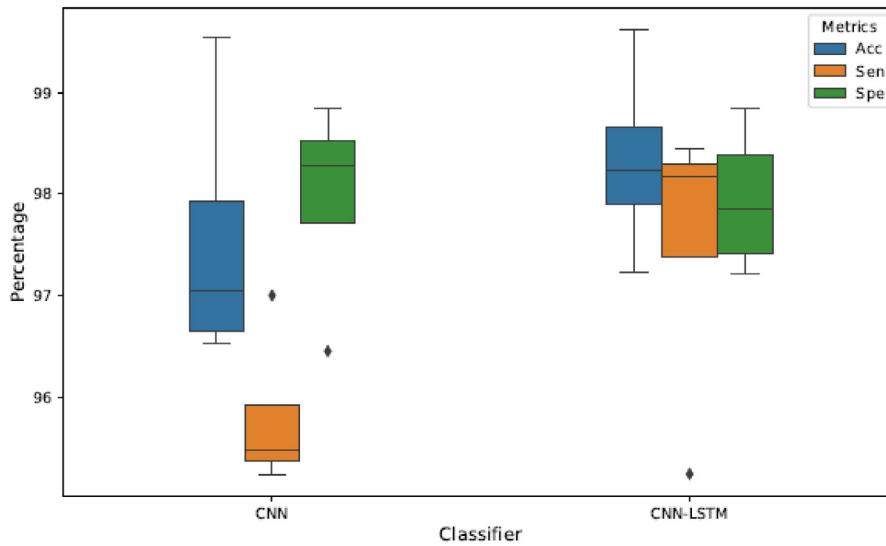


Fig. 16. Box plot of the key performance metrics for CNN and CNN-LSTM..

regularization technique combined with dropout significantly reduces the overfitting issue. To provide the statistical illustration of the classifier models, the box plot of CNN and CNN-LSTM is shown in Fig. 16. We can note that the performance of the hybrid CNN-LSTM is significantly better than that of the CNN classifier.

5.4. Performance comparison with existing techniques

For validation, the classification performance of the proposed framework with those of the other existing methods, which have utilized gait analysis for PD diagnosis, is presented in Table 7. The significant

Table 7

Performance comparison of proposed and existing approaches.

References	Classification	Classifier	Acc (%)	Spe (%)	Sen (%)
Ashour et al. (2020)	Two class	LSTM	83.90	84.90	88.74
El Maachi et al. (2020)	Multi-class	CNN	85.22	85.28	87.29
Veeraragavan et al. (2020)	Multi-class	ANN	87.11	90.48	67.89
Oktay and Kocer (2020)	Two class	LSTM	90.01	87.00	95.23
Khoury et al. (2019)	Two class	RF	90.90	88.40	85.36
Abdulhay et al. (2018)	Two class	SVM	92.68	92.22	96.14
Zhao et al. (2018a)	Multi-class	LSTM	93.19	96.13	94.80
Balaji et al. (2021b)	Multi-class	LSTM	96.61	96.19	98.10
Proposed	Multi-class	CNN-LSTM	98.32	97.68	98.29

advantages of the proposed approach over other competing approaches are as follows. In Refs. Khouri et al. (2019), Abdulhay et al. (2018), Ashour et al. (2020) and Oktay and Kocer (2020) only the binary classification problem was addressed using the machine learning and deep learning algorithms, respectively. However, our work addresses the multi-class classification problem and classifies the stages of PD based on UPDRS and H&Y scale. Similarly, compared to the recently proposed deep learning techniques in Refs. Maiti et al. (2017), Balaji et al. (2021b), Zhao et al. (2018b) and Veeraragavan et al. (2020) that have addressed multi-class PD classification problem, the proposed approach offers better accuracy, sensitivity and specificity. Specifically, compared to the classification performance of LSTM classifier for PD severity put forward by Balaji et al. (2021b), this work achieves a 1.7% and 2.1% improvements in accuracy and sensitivity, respectively. Hence, the proposed CNN-LSTM framework is highly accurate and robust in predicting the stages of PD by automatically extracting the prominent features from the IMFs of gait cycle.

5.4.1. Limitations of the proposed approach

In spite of the considerable improvement in the classifier performance, two major limitations of the proposed approach which require further investigations are as follows. Firstly, this technique can be applied only for people who can walk and have mild to moderate gait impairments. Hence, as a future work, along with the evaluation of motor symptoms, the assessment of nonmotor symptoms can also be integrated for diagnosing subjects who either in need of assistance while walking or wheel chair bounded. Secondly, CNN-LSTM network requires considerable effort by trial and error to find the optimal hyper parameters. Hence, hyper parameter selection can be formulated as an optimization problem and the efficacy of evolutionary computation algorithms to solve the optimal hyper parameter selection problem can be explored. Moreover, the computational time of the proposed approach to train the DNN model is approximately 3 h. Hence, as a future work, we also plan to devise a light-weight CNN-LSTM architecture which could potentially minimize the computational load and time.

6. Conclusions

This paper has put forward a gait impairment based PD stage classification framework using the CNN-LSTM classifier. As the gait variability is the prominent symptoms of PD, this study has considered the VGRF time series data for three walking tests from Physionet database to predict the severity of PD. Unlike most of the previous studies which have utilized all the VGRF channels for extracting the potential gait biomarkers to differentiate between the healthy controls and PD subjects, we present a framework based on FAV to select the optimal number of VGRF channels for reducing the computational load on the deep learning classifier model. Moreover, the VGRF signals from prominent sensors have been decomposed using the EMD technique in order to denoise and decompose the VGRF signals into IMFs. Through power spectral analysis the dominant IMFs have been identified to extract the deep features using CNN. Moreover, the extracted CNN feature vectors are utilized in LSTM to classify the stages of PD according to H&Y scale. The L2 regularization along with dropout techniques has

been employed to address the data overfitting problem. Furthermore, Adam optimizer that requires less memory and minimal tuning has been used to solve the stochastic objective function. The comprehensive analysis of the classification performance of the proposed CNN-LSTM model against the state-of-the-art techniques highlights that the proposed technique is more accurate with less computational burden. Hence, the proposed gait classification framework can aid the physicians as a decision support system to identify the severity of PD by automatically extracting the potential gait biomarkers through deep learning classifier model.

CRediT authorship contribution statement

B. Vidya: Conceptualization, Methodology, Visualization, Data curation, Software, Writing – original draft. **Sasikumar P.:** Resources, Supervision, Project administration, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Mei, J., Desrosiers, C., Frasnelli, J., 2021. Machine learning for the diagnosis of Parkinson's disease: A review of literature. *Front. Aging Neurosci.* 13, 184.
- Camps, J., Sama, A., Martin, M., Rodriguez-Martin, D., Perez-Lopez, C., Arostegui, J.M.M., Cabestany, J., Catala, A., Alcaine, S., Mestre, B., et al., 2018. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowl.-Based Syst.* 139, 119–131.
- Goyal, J., Khandnor, P., Aseri, T.C., Classification, prediction, 2020. And monitoring of Parkinson's disease using computer assisted technologies: A comparative analysis. *Eng. Appl. Artif. Intell.* 96, 103955.
- Ricciardi, C., Amboni, M., De Santis, C., Impronta, G., Volpe, G., Iuppariello, L., Ricciardelli, G., D'Addio, G., Vitale, C., Barone, P., et al., 2019. Using gait analysis' parameters to classify Parkinsonism: A data mining approach. *Comput. Methods Programs Biomed.* 180, 105033.
- Maiti, P., Manna, J., Dunbar, G.L., 2017. Current understanding of the molecular mechanisms in Parkinson's disease: Targets for potential treatments. *Transl. Neurodegener.* 6 (1), 1–35.
- Zhao, J., Mao, X., Chen, L., 2019. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomed. Signal Process. Control* 47, 312–323.
- Wahid, F., Begg, R.K., Hass, C.J., Halgamuge, S., Ackland, D.C., 2015. Classification of Parkinson's disease gait using spatial-temporal gait features. *IEEE J. Biomed. Health Inf.* 19 (6), 1794–1802.
- Kamran, I., Naz, S., Razzak, I., Imran, M., 2021. Handwriting dynamics assessment using deep neural network for early identification of parkinson's disease. *Future Gener. Comput. Syst.* 117, 234–244.
- Farashi, S., 2020. Distinguishing between parkinson's disease patients and healthy individuals using a comprehensive set of time, frequency and time-frequency features extracted from vertical ground reaction force data. *Biomed. Signal Process. Control* 62, 102132.
- Cantürk, İ., 2021. A computerized method to assess Parkinson's disease severity from gait variability based on gender. *Biomed. Signal Process. Control* 66, 102497.
- Balaji, E., Brindha, D., Balakrishnan, R., 2020. Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease. *Appl. Soft Comput.* 94, 106494.

- Khoury, N., Attal, F., Amirat, Y., Oukhellou, L., Mohammed, S., 2019. Data-driven based approach to aid Parkinson's disease diagnosis. *Sensors* 19 (2), 242.
- Alam, M.N., Garg, A., Munia, T.T.K., Fazel-Rezai, R., Tavakolian, K., 2017. Vertical ground reaction force marker for Parkinson's disease. *PLoS One* 12 (5), e0175951.
- Joshi, D., Khajuria, A., Joshi, P., 2017. An automatic non-invasive method for Parkinson's disease classification. *Comput. Methods Programs Biomed.* 145, 135–145.
- Li, Q., Li, L., Wang, W., Li, Q., Zhong, J., 2020. A comprehensive exploration of semantic relation extraction via pretrained CNNs. *Knowl.-Based Syst.* 194, 105488.
- El Maachi, I., Bilodeau, G.-A., Bouachir, W., 2020. Deep 1D-convnet for accurate Parkinson disease detection and severity prediction from gait. *Expert Syst. Appl.* 143, 113075.
- Balaji, E., Brindha, D., Elumalai, V.K., Vikrama, R., 2021b. Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network. *Appl. Soft Comput.* 108, 107463.
- Zhao, A., Qi, L., Li, J., Dong, J., Yu, H., 2018b. A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data. *Neurocomputing* 315, 1–8.
- Zhang, K., Wang, W., Lv, Z., Fan, Y., Song, Y., 2021. Computer vision detection of foreign objects in coal processing using attention CNN. *Eng. Appl. Artif. Intell.* 102, 104242.
- Wu, Y., Krishnan, S., 2009. Statistical analysis of gait rhythm in patients with Parkinson's disease. *IEEE Trans. Neural Syst. Rehabil. Eng.* 18 (2), 150–158.
- Lee, S.-H., Lim, J.S., 2012. Parkinson's disease classification using gait characteristics and wavelet-based feature extraction. *Expert Syst. Appl.* 39 (8), 7338–7344.
- Perumal, S.V., Sankar, R., 2016. Gait and tremor assessment for patients with Parkinson's disease using wearable sensors. *ICT Express* 2 (4), 168–174.
- Prabhu, P., Karunakar, A.K., Anitha, H., Pradhan, N., 2018. Classification of gait signals into different neurodegenerative diseases using statistical analysis and recurrence quantification analysis. *Pattern Recognit. Lett.*
- Yurdakul, O.C., Subathra, M., George, S.T., 2020. Detection of Parkinson's disease from gait using neighborhood representation local binary patterns. *Biomed. Signal Process. Control* 62, 102070.
- Ashour, A.S., El-Attar, A., Dey, N., Abd El-Kader, H., Abd El-Naby, M.M., 2020. Long short term memory based patient-dependent model for FOG detection in Parkinson's disease. *Pattern Recognit. Lett.* 131, 23–29.
- Oktay, A.B., Kocer, A., 2020. Differential diagnosis of Parkinson and essential tremor with convolutional LSTM networks. *Biomed. Signal Process. Control* 56, 101683.
2021. Gait in parkinson's disease. Available online: <https://www.physionet.org/content/gaitpdb/1.0.0/> (accessed on 20 January 2021).
- Yogev, G., Giladi, N., Peretz, C., Springer, S., Simon, E.S., Hausdorff, J.M., 2005. Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding? *Eur. J. Neurosci.* 22 (5), 1248–1256.
- Hausdorff, J.M., Lowenthal, J., Herman, T., Gruendlinger, L., Peretz, C., Giladi, N., 2007. Rhythmic auditory stimulation modulates gait variability in Parkinson's disease. *Eur. J. Neurosci.* 26 (8), 2369–2375.
- Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., Hausdorff, J.M., 2005. Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease. *Mov. Disord.: Off. J. Mov. Disord. Soc.* 20 (9), 1109–1114.
- Balaji, E., Brindha, D., Elumalai, V.K., Umesh, K., 2021a. Data-driven gait analysis for diagnosis and severity rating of Parkinson's disease. *Med. Eng. Phys.* 91, 54–64.
- Arivazhagan, S., Amrutha, E., Jebarani, W.S.L., 2021. Universal steganalysis of spatial content-independent and content-adaptive steganographic algorithms using normalized feature derived from empirical mode decomposed components. *Signal Process., Image Commun.* 101, 1–11.
- Vican, I., Kreković, G., Jambrošić, K., 2021. Can empirical mode decomposition improve heartbeat detection in fetal phonocardiography signals? *Comput. Methods Programs Biomed.* 203, 106038.
- Sharma, G., Parashar, A., Joshi, A.M., 2021. Dephn: a novel hybrid neural network for electroencephalogram (EEG)-based screening of depression. *Biomed. Signal Process. Control* 66, 102393.
- Kim, T.-Y., Cho, S.-B., 2019. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* 182, 72–81.
- Chen, C., Hua, Z., Zhang, R., Liu, G., Wen, W., 2020. Automated arrhythmia classification based on a combination network of CNN and LSTM. *Biomed. Signal Process. Control* 57, 101819.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Hussein, R., Palangi, H., Ward, R.K., Wang, Z.J., 2019. Optimized deep neural network architecture for robust detection of epileptic seizures using eeg signals. *Clin. Neurophysiol.* 130 (1), 25–37.
- Abdulhay, E., Arunkumar, N., Narasimhan, K., Vellaippalan, E., Venkatraman, V., 2018. Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease. *Future Gener. Comput. Syst.* 83, 366–373.
- Veeraragavan, S., Gopalai, A.A., Gouwanda, D., Ahmad, S.A., 2020. Parkinson's disease diagnosis and severity assessment using ground reaction forces and neural networks. *Front. Physiol.* 11.
- Zhao, A., Qi, L., Dong, J., Yu, H., 2018a. Dual channel LSTM based multi-feature extraction in gait for diagnosis of neurodegenerative diseases. *Knowl.-Based Syst.* 145, 91–97.