



# Novel machine learning-based hybrid strategy for severity assessment of Parkinson's disorders

Preeti Khara<sup>1,2</sup> · Neelesh Kumar<sup>1,2</sup>

Received: 16 April 2021 / Accepted: 21 January 2022 / Published online: 4 February 2022  
© International Federation for Medical and Biological Engineering 2022

## Abstract

Parkinson's disease (PD) severity assessment in clinical settings largely depends on expertise level of clinicians which have inherent limitations and non-uniformity. Instrumented gait analysis plays a significant role in disease diagnosis and management. However, these are agonized from data dispersion due to demography, anthropometry, and self-selected walking speed of an individual. This research work aims to develop computationally efficient hybrid strategy using normalized features for PD severity evaluation. The relevance of each considered gait feature in demonstrating the outcomes is explained through feature importance and partial dependence plot (PDP) to build substantial insight for clinical needs. Gait, a biomarker, is used to access human mobility by utilizing vertical ground reaction force (VGRF) data of 72 healthy and 93 Parkinson's individuals. A multi-variate normalization approach identifies gait differences between PD and non-PD. The proposed hybrid model used is able to detect PD with accuracy of 99.39% and 99.9%, and its severity assessment based on MDS-UPDRS-III shows coefficient of determination ( $R^2$ ) as 97% and 98.7% using leave-one-out cross-validation (CV) and tenfold CV respectively. The significant features suitable for clinical implications are reported. Moreover, normalized gait parameters supplement capability to compare individuals with diverse physical properties, resulting in assistive system for evaluation of PD severity.

**Keywords** Gait · Machine learning · Parkinson's severity · Normalization · UPDRS

## 1 Introduction

Quantitative gait assessment is significant in Parkinson's disease (PD) diagnosis and management. Despite of advances in medical and healthcare, gait impairments worsen with progression of the disease, leading to fall-risk, loss of independence, and degradation in quality life [1]. The disease diagnosis can take several years due to inherent variability in sign, symptoms, and frequency based on individuals' medical records and neurological analysis [2]. The quantification of disease severity is carried out using various clinical scales [3, 4]. Out of which, unified Parkinson disease rating scale (UPDRS) is the most commonly used scale for PD assessment. The analysis based on the scale [4] is a collection of self-assessed questionnaires, and ratings through

visual observation by experts to evaluate behavioral characteristics, daily activities of living, motor symptoms, and complications of therapy. The evaluation based on expert opinion leads to bias adding unreliability and low efficiency [5]. Moreover, the current development in wearable sensors allows monitoring motor symptoms, which in turn has made gait analysis possible in unconstrained environment but still lacks decision-making. Thus, to make inferences predictive models capable of processing high-dimensional complex data is required.

Machine learning (ML) techniques have gained popularity in recognizing activities of daily living. They are widely used in gait analysis for classifying individuals with neurological disorders [6, 7], quantifying pre and post-intervention outcomes [8]. The potential application areas to use such techniques for PD diagnosis include speech signal [9], handwriting variations [10], and gait variability [11–13]. Particularly, the variation in gait patterns can act as a significant biomarker to diagnose PD as well as to determine progression of disease [14]. Rather than focusing on statistical time or frequency domain features that lack clinical meaning, this study investigates to extract clinically relevant features from raw VGRF.

✉ Neelesh Kumar  
neel5278@csio.res.in

<sup>1</sup> Academy of Scientific and Innovative Research,  
Ghaziabad 201002, India

<sup>2</sup> Biomedical Instrumentation Unit, CSIR-CSIO,  
Chandigarh 160030, India

The extracted features based on domain knowledge, and related to pace, rhythm, variability in spatial–temporal domain, force variation, and temporophasic represent a greater potential for disease classification. Nevertheless, the spatio-temporal features [15] are affected by height, body mass, gender, and by aging [16] which limits the degree to which pathological traits can be reliably distinguished. However, such parameters are also influenced to a great extent by walking speed [17] of an individual. Also, the force-related features are affected by body mass [18] leading to unpredictability in pathology detection.

Therefore, to reduce data dispersion, it is effective to minimize the effect of inter-subject variations to improve severity assessment. The data normalization using dimensionless equations accounts for subjects' body mass and height using inverted pendulum analogy [19]. However, such techniques do not account for variation in multiple characteristics. A study [15] proposed a normalization approach to account for such differences. The authors hypothesized that the inclusion of this approach would significantly improve severity assessment of PD. Thus, the study aims to extract clinically relevant features from foot sensor data and development of hybrid strategy using ML for PD severity assessment based on motor disorders. The main contribution of the research work includes:

- To analyze normalization approach for correcting individual differences in extracted features due to individuals' demography, anthropometry, and walking speed between non-PD and PD individuals.
- To determine PD severity rating through developed novel hybrid strategy. The first stage is used to distinguish a PD individual from non-PD, and thereby accessing their severity in the second stage.
- To compare the performance of the hybrid model with traditional methods, including evaluation of model through various performance metrics.
- To propose a relevant feature set using interpretability techniques for disease severity assessment to build clinical insight.

The paper is structured as follows: Section 2 discusses literature related to PD detection. Section 3 briefs about the dataset, data normalization, and proposed model. The results and findings together with various evaluation metrics are described in Section 4. Section 5 presents the comparative analysis with existing work. Finally, Section 6 concludes by highlighting significant outcomes of the research work.

## 2 Background work

The existing studies employ kinematic and/or kinetic signal to extract both low- and high-level features. Some of the features included in previous work comprise mean, median,

maximum, and minimum value with its index, range [15, 16], kurtosis, skewness, entropy, root mean square, inter-quartile range, harmonic mean, phase, energy, and power [20]. Moreover, the high-level gait features are more informative, interpretable, and can be computed using more sophisticated methods. Previous studies rely on statistical approaches [21–23]; undoubtedly, such methods provide transparency in determining discriminative power and contribution of each individual feature. But converting their results into automated classification results in thresholding which can lead to bias and classification errors.

However, ML techniques are more powerful to draw such inferences. The well-explored learning models reported in literature for PD classification include support vector machine (SVM) [24, 25], neural networks (NNs) [26], k-nearest neighbors (kNN) [27, 28], Bayes network (BN) [26], naïve Bayes (NB) [26], logistic regression (LR) [26], random forest (RF) [29], decision trees (DT) [29], Gaussian mixture model (GMM) [28], and 1D-Convnets [12]. Such techniques are capable of processing high-dimensional, complex, and multi-variate feature set possessing non-linearity and potentially leading to superior performance. Most of these models consist of statistical, time, or frequency domain features and aid in detecting PD. But high-level gait parameters are more valuable to build clinical insight for disease classification. Also, many existing studies are related to objective assessment that reveals only PD abnormality. However, few studies [11–13, 20, 30] quantify severity of disease according to Hoehn and Yahr (H&Y) scale, and studies [12, 20] aim to quantify disease severity based on UPDRS using raw foot sensor signal and statistical measures.

The studies worth mentioning are presented here; in [20], the authors developed a computational solution to analyze PD and its severity based on regression analysis using statistical time and frequency domain features, thereby, obtaining high inferences. Another study [12] used 1D-convnets having high computational burden to determine PD severity based on UPDRS. They classify PD subjects into categorical form by dividing clinical scores into five classes, leading to low classification accuracies. Thus, it can be concluded from literature screening that the classification accuracy reported for severity assessment still needs high inferences and clinical implications using clinical gait parameters. There is also a need to devise a computationally efficient methodology [30] for studying PD severity assessment more precisely. However, UPDRS scale constitutes measures of non-motor symptoms along with motor components which are difficult to quantify using instrumentation analysis. Also, quantification of disease based on UPDRSm (MDS-UPDRS-III) is advantageous than H&Y scale due to its capability to inherit the measures provided by later scale.

Considering the limitations of existing work, the significant features for disease classification are extracted,

normalized, and used as an input to the hybrid model. The hybrid model provides PD severity assessment based on a clinical scale (UPDRSm) for motor disorders. The process flow highlighting the progress in existing work till the proposed methodology is demonstrated in Fig. 1. The details of most relevant studies considering PD detection are outlined in Table 1. The merits and demerits of included studies are also mentioned.

### 3 Methodology

In this section, details about considered dataset, data pre-processing, feature extraction, data normalization, and methodology for PD severity assessment is presented. To reduce inter-subject gait variability, and to provide quantitative assessment of PD symptoms, a novel hybrid strategy is developed in this work. The main significance of the proposed approach is to determine the exact value of PD severity rating based on motor disorders using instrumented gait analysis rather than obtaining the categorical results that define the severity levels. Thus, the proposed model can benefit the clinicians with precise prognostic outcomes rather than binary diagnostic decision. Further, to improve severity prediction normalization is carried out that excludes dependency due to demography, anthropometry, and self-selected walking speed of an individual. Finally, the most significant features for PD and its severity evaluation are identified.

### 3.1 Dataset

An open access dataset of gait in Parkinson’s disease collected using foot insole sensors (Ultraflex Computer Dynography, Infotronic Inc.) is taken from Physionet [34]. The database includes 306 gait recordings (19,320 gait cycles) from 72 healthy and 93 idiopathic Parkinson’s. This dataset collected at the movement disorder unit of Tele-Aviv Sourasky Medical Centre, Laboratory for Gait and Neurodynamics was established by three different research groups. The study by Frenkel-Toledo [21] includes gait time series for normal ground walk and walk on motorized treadmill. Another group [22] includes walk with and without rhythmic auditory stimulation (RAS) at a comfortable pace. Finally, the third study [23] comprises leveled ground walk at self-selected speed under conditions such as usual walking and dual tasking. Each recording consists of VGRF data from 16 sensors, in which each insole contains 8 sensors and rest two accounts for total VGRF signal from both foot sensors. The data acquisition protocol constitutes walk of 2 min at self-selected speed on leveled surface. The VGRF signals are sampled at 100 Hz and measured in Newtons. Thus, to identify the PD severity, in this work all datasets are taken together because the gait dynamics through force sensor data can unveil underlying locomotor disability which aids in quantifying stride-to-stride variability. Along with demographic measures, dataset includes details of mobility (Timed Up and Go Test (TUG)) and severity measures based on clinical scales (H&Y, UPDRS, UPDRSm). Table 2 highlights the demographic detail of the included subjects. The acquisition system included a pair of shoes equipped with

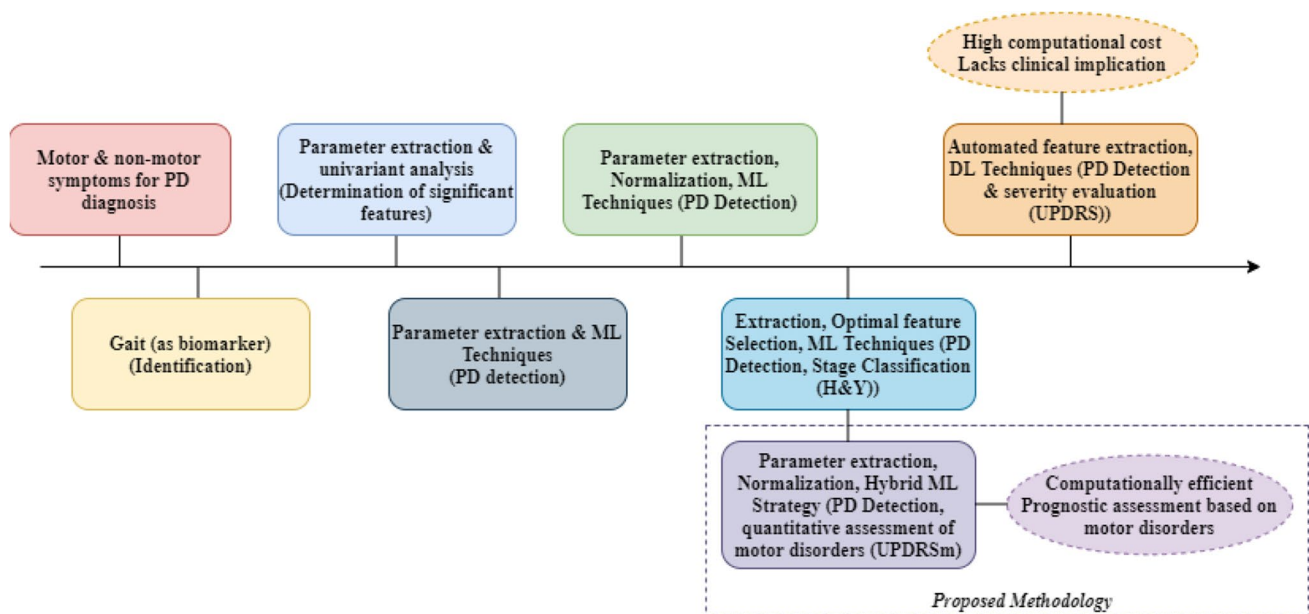


Fig. 1 Process flow highlighting the progress in existing work till proposed methodology

**Table 1** Literature survey on PD and its severity assessment using gait biomarkers

Study	Objectives	Methodology	Cross-validation (CV)	Merits	Demerits
(Zhang et al., 2013) [25]	Sparse representation of gait signal to differentiate PD from non-PD	ML techniques (SVM, D-KSVD, LC-KSVD)	tenfold	Achieved classification accuracy (83.44%)	Lacks assessment of PD severity
(Alkhatib et al., 2015) [27]	Statistical feature extraction and feature selection by evaluation on ROC curves to determine optimal feature set for achieving accurate prediction about PD	Feature selection (evaluation on ROC curves) ML technique (kNN)	-	Able to attain prediction accuracy as 83% Sensors at inner arch of insole possess better discriminating power for classification	Statistical features PD severity not identified
(Wahid et al., 2015) [31]	PD detection using spatio-temporal features	ML techniques (KFD, NB, kNN, SVM, RF)	tenfold	Multiple regression normalized features predict PD with 92.6% accuracy using RF	Lacks PD severity prediction Less no. of subjects
(Perumal et al., 2016) [24]	Gait feature extraction to differentiate a PD from non-PD Frequency domain features for tremor analysis	Statistical analysis (ANOVA) ML techniques (LDA)	fivefold	Able to classify PD subjects with 86.9% accuracy, 0.72 as sensitivity, and 0.81 as specificity	Severity of PD not identified
(Ertugrul et al., 2016) [26]	Statistical features both in time and frequency domain using shifted ID-LBP were extracted to identify PD subject	Feature selection (RandomF) ML models (BN, NB, LR, MLP, DT, RF, FT)	tenfold	Using all 108 features able to achieve classification accuracy (88.88%), sensitivity (0.889), specificity (0.822) with MLP classifier, but using selected 21 features and RF classifier accuracy was 87.58%, sensitivity (0.876), and specificity (0.767)	PD severity not accessed Feature selection reduces dimensionality of the signal features but at the same time specificity of the model also reduces
(Zeng et al., 2016) [32]	Deterministic learning approach to classify PD from non-PD individuals	ML techniques (RBF-NN)	twofold fivefold leave-one-out	Able to achieve high inference accuracies (96.39%), sensitivity (0.96), and specificity (0.95)	Lacks inclusion of age, gender, medication, and other gait dynamics effects that could aid in improving inferences and strengthening clinical implications
(Alam et al., 2017) [29]	Feature extraction (gait + statistical features) and selection of best suited features to differentiate PD individuals	Feature selection (FFS, MRMR, mutual information) ML techniques (SVM, kNN, DT, RF)	Leave-one-out	PD individuals were identified with accuracies ranging between 85.1% and 95.7% using different classifiers	To achieve more reliability more clinical features are required Lacks PD severity assessment
(Zhao et al., 2018) [30]	Automated spatio-temporal feature learning using deep learning architectures	A hybrid fusion model comprising LSTM and CNN	tenfold	High inference accuracies for PD and its severity rating based on H&Y scale	High computational cost Lacks clinical insights for medical implications
(Prashanth et al., 2018) [33]	Stage estimation using MDS-UPDRS and H&Y to determine stages of the disease	Feature importance (RF) ML techniques (OLR, SVM, AdaBoost, RUS Boost)	tenfold	Large no. of subjects High accuracy (AdaBoost: 97.46%)	Lacks determination of feature importance between PD stages Assessment of severity on categorical scale (normal, early, or moderate)

**Table 1** (continued)

Study	Objectives	Methodology	Cross-validation (CV)	Merits	Demerits
(Maachi et al., 2019) [12]	Raw VGRF signals from both foot sensors to determine PD and its severity using UPDRS	ID-Convnets	tenfold	High accuracies for PD detection (98.7%) as compared to its severity assessment (85.3%)	Low inference outcomes for PD severity and being categorical divided as 5 classes High computational cost
(Balaji et al., 2020) [11]	Early detection and staging of PD with clinical scale H&Y using statistical and kinematic analysis	ML techniques (DT, SVM, EC, BN)	tenfold	High classification accuracies (99.4%) using both features	Lacks consideration of physical characteristics like height, weight, and by aging
(Khera et al., 2021) [13]	To access PD and its staging based on H&Y scale and to determine the influence of age and gender for PD detection	Correlation analysis Feature selection (RFE) ML techniques (SVM, kNN, DT, RF)	tenfold	High inference accuracies (98.50%) reported for age-gender based severity prediction model	Lacks inclusion of other parameters like height, body weight, and walking speed which can lead to strengthen the prediction outcomes

Abbreviations: *SVD* singular value decomposition, *ROC* receiver operating characteristics, *KFD* Kernel Fisher discriminant, *LDA* linear discriminant analysis, *LBP* local binary patterns, *FT* functional tree, *FFS* forward feature selection, *MRRM* minimum redundancy maximum relevance, *LSTM* long short-term memory, *CNN* convolutional neural network, *OLR* ordinal logistic regression, *EC* ensemble classifier, *RFE* recursive feature elimination

sensory module and a recording unit (19\*14\*4.5 cm; 1.5 kg) attached to the waist. Figure 2 illustrates the position of each sensor in both feet to capture VGRF recordings, and relative position of each sensor with origin (0, 0) between both legs towards positive y-axis is given in Table 3.

### 3.2 Pre-processing and feature extraction

The raw sensor data is filtered and segmentation is carried out to extract gait cycles [13]. Further, to remove startups and termination effects 20 s of data from both initiation and termination of each gait cycle is removed. In total 16 gait characteristics are derived and categorized into broad independent domains. The progression of PD can be accessed through gait variations as the disease affects the walking pattern of the individual. Therefore, a study comprising disease significant features helps to understand gait disorder of PD subjects and acts as a significant biomarker. The features extracted using raw sensor data comprise stance-interval, swing-interval, step-time, stride-length, stride-interval, step-length, center of pressure in x and y directions, heel-strike force, toe-off force, double limb support, percent of temporal features with respect to gait cycle, and ratio of swing to stance interval. Table 4 comprehends various features extracted using VGRF, and their distribution is plotted in Fig. 3.

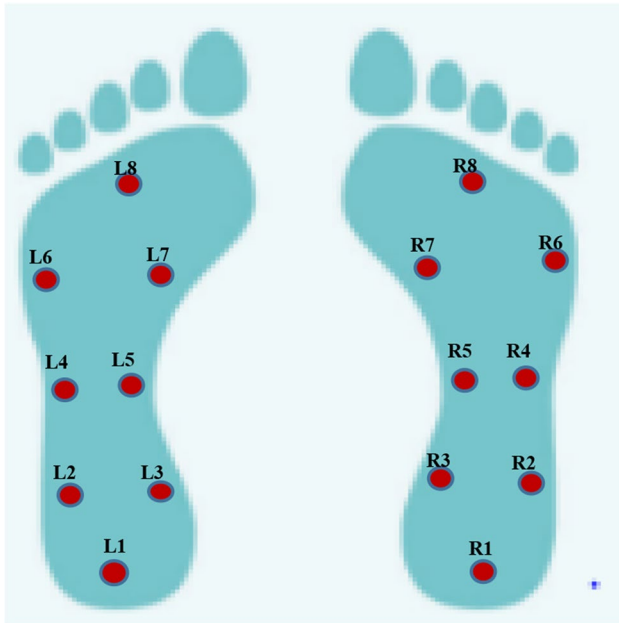
### 3.3 Data normalization

Nevertheless, gait data accounts for variation due to demography, anthropometry, and self-selected walking speed of an individual. In this case classification results would be biased leading to decrease in performance and more prone to errors for determining disease severity. So, to gain efficient severity prediction and to exclude all such dependencies from the features, multi-variate regression model selection followed by backward elimination is implemented [15]. The statistical assumptions required for each independent variable including linearity, normality, and homoscedasticity are met. First, correlation between various independent variables (such as gender (G), age (A), height (H), weight (W), and self-selected walking speed (S)) is determined using Spearman’s rank correlation. To address the issue of multicollinearity, variance inflation factor (VIF) of each variable is computed. If VIF exceeds 5, then variance decomposition values over 0.5 and condition index exceeding 30 are indicators of multicollinearity. So, such parameters are not to be included in the final model. After selection of independent variables, regression model for each dependent variable (gait features as listed in Table 4) is constructed by removing least significant independent variable ( $p < 0.001$ ). The process continues until only significant variables remain in the model. To avoid overfitting, the tenfold cross validation root mean



**Table 2** Demographic details of gait in Parkinson's dataset

S. no	Category of subject	No. of subjects	Age (in years)	Gender		Walking speed (m/s)	Mobility measure (TUG)
				Male	Female		
1	Healthy (HC)	72	63.65 ± 8.58	40	32	1.24 ± 0.159	9.29 ± 1.57
2	Parkinson's (PD)	93	66.30 ± 9.45	58	35	1.033 ± 0.205	12.05 ± 3.94

**Fig. 2** Sensor location inside insoles**Table 3** Relative positioning of sensors

Sensors	X axis (mm)	Y axis (mm)
L1	-500	-800
L2	-700	-400
L3	-300	-400
L4	-700	0
L5	-300	0
L6	-700	400
L7	-300	400
L8	-500	800
R1	500	-800
R2	700	-400
R3	300	-400
R4	700	0
R5	300	0
R6	700	400
R7	300	400
R8	500	800

square error (RMSE-CV) is computed and compared with model's RMSE. Thereafter, normalization of each parameter is carried out using Eqs. (1)–(3).

$$z_i = \alpha_o + \sum_{j=1}^n \alpha_j x_{ij} + \varepsilon_i \quad (1)$$

where,  $z_i$  represents dependent variables for the  $i^{\text{th}}$  observation;  $x_{ij}$  denotes independent variables corresponding to  $i^{\text{th}}$  observation and  $\alpha_j$  are the coefficients of model for particular independent variable obtained in the final model;  $\alpha_o$  represents intercept term of regression model; independent residual error is represented as  $\varepsilon_i \sim N(0, \sigma^2)$ . All spatio-temporal features are normalized using this approach, and model coefficients can be determined using Eq. (1).

$$z_n = \frac{z_i}{z_{i\text{-pred}}} \quad (2)$$

where,  $z_n$  is the normalized value of gait feature computed by dividing the raw value with model predicted value shown as

$$z_i = z_{i\text{-pred}} + \varepsilon_{i\text{-pred}} \quad (3)$$

The model's ability is evaluated using Spearman's rank order correlation with raw and normalized data. The reduction in data dispersion is determined using coefficient of variation with 95% confidence level (CL) and standard error (SE) [35]. The statistical significance of outcomes is evaluated ( $p < 0.05$ ) using Student  $t$  test.

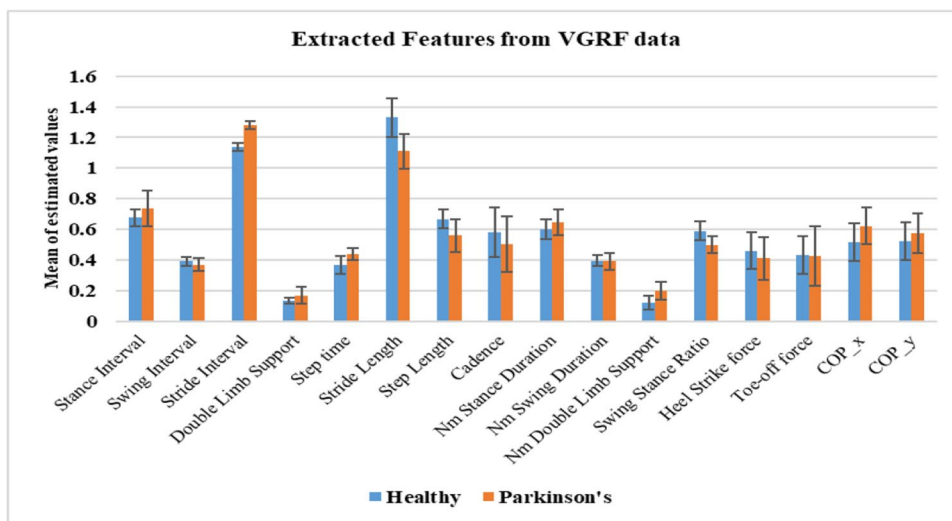
### 3.4 Hybrid strategy

A two-stage ML hybrid strategy is designed for assessment of PD and its severity. The first stage comprises DT classifier that predicts PD gait; thereafter, ensemble regressors (ER) are tuned to determine severity of PD. The classifier internal parameters are tuned using grid search and cross-validated with tenfold and leave-one-out (LOOCV). In DT classifier, Gini's diversity index is chosen as split criterion consisting of maximum splits as 100. The DT assigns the data in the form of nodes and branches and thus is able to determine the non-linearity between features and target response of the system. It results in better generalization on unseen data

**Table 4** List of extracted features with their type and definition

Type	Gait features	Definition
Time	Stance-interval (stn-dur)	Duration for which one foot is in contact with ground
	Swing-interval (sw-dur)	Duration for which one foot is off the ground
	Stride-interval (str-dur)	Duration between two consecutive events of the same foot
	Double limb support (DS-dur)	Duration of bilateral foot contact
	Step-time (stp-time)	Duration between initial contact of one foot to initial contact of contralateral foot
Length	Stride-length (str-length)	Total distance between successive ground contacts of same foot
	Step-length (stp-length)	Distance between initial contact of one foot to initial contact of contralateral foot
Frequency	Cadence	Number of steps taken per minute
Temporophasic	Normalized (Nm) stance duration (nm-stn-dur)	Stance interval measured with respect to gait cycle time or stride time, i.e., stn-dur/str-dur
	Normalized (Nm) swing duration (nm-sw-dur)	Swing interval measured with respect to gait cycle time or stride time, i.e., sw-dur/str-dur
	Normalized (Nm) double limb support (nm-DS-dur)	Double limb support measured with respect to gait cycle time or stride time, i.e., DS-dur/str-dur
	Swing stance ratio (SSR)	Ratio of swing to stance intervals
Force	Heel-strike force (HS-force)	Mean of sensor values beneath the heel for initial 5% sample points in stance duration of total gait cycle
	Toe-off force (TO-force)	Mean of sensor values beneath the toe for final 5% sample points in stance duration of total gait cycle
	Centre of pressure (x,y) (COP_x, COP_y)	Total sum of pressure field acting on a body causing force to act on ground

**Fig. 3** Distribution of extracted features using VGRF data



[36]. In order to predict PD severity, the second stage of the model is trained using collection of tree-based ER model built using bagged approach with 30 as number of learners containing leaf size of 8 and 0.1 as learning rate. The proposed approach provides an exact value of PD severity rating based on motor disorders using instrumented gait analysis rather than obtaining the categorical outcomes that define severity levels.

The performance metrics considered for evaluating the efficacy of hybrid strategy are accuracy (Acc.) which is described as the proportion of correct predictions to the total number of predictions, specificity (Sp.) that measures true negatives to actual negatives that is correctly identifying individuals without a disease, and sensitivity (Sen.) or recall that determines the number of true positives to total number of actual positives which helps in identifying the individual with a disease condition. They are good indicators

for identifying the specific condition of an individual. The positive predicted values are termed as precision (prec.). It conveys how many of diagnostic test positives are true positives; the higher value indicates superior performance. Finally, F1-score defines the harmonic mean of both prec. and Sen./recall (Eq. 4). The variability score  $\alpha = 1$  is chosen as default for equal importance of both variables.

$$F1 - score(\%) = (1 + \alpha^2) \frac{prec. * Sen.}{(\alpha^2 * prec.) + Sen.} * 100 \tag{4}$$

The mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination ( $R^2$ ) are presented in Eqs. (5–7).

$$MAE = \frac{1}{n} \sum_{j=1}^n | (y_{pred_j} - y_j) | \tag{5}$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^n (y_{pred_j} - y_j)^2}{n}} \tag{6}$$

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{j=1}^n (y_j - y_{pred_j})^2}{\sum_{j=1}^n (y_j - \bar{y}_j)^2} \tag{7}$$

where,  $n, y_j, y_{pred_j}$  represent total number of strides, actual value, and predicted value by the model.  $\bar{y}_j$  is the average of  $[y_1, y_2, \dots, y_n]$ , thus, a model where predicted values exactly match the observed value. The sum of square of residuals (SSR) is zero and  $R^2 = 1$ . However, variance of the data is determined by total sum of squares (SST).

### 3.5 Interpretability in machine learning

ML models, however, are widely used in the medical domain, but explaining the cause of prediction can assist clinicians in treatment planning. To estimate the feature importance in this work, out-of-bag feature ensemble method is used to arbitrarily permute the value of extracted features and thereby examining their effect on resulting error bias. After defining the most relevant features, partial dependence plot (PDP) [37–39] is used to determine the marginal contribution of each considered gait parameter to predicted UPDRSm of hybrid model allowing better understanding of the model. The individual conditional expectation (ICE) indicates the dependence of prediction on specific feature for each patient, while PDP represents averaged contribution of the predicted values on model outcome. Thus, keeping all the other features same, partial

dependency of prediction is determined based on changes in prediction scores (y-axis) by replacing the values of considered gait feature (x-axis).

#### 3.5.1 Mathematical model

The partial dependency for severity prediction in PD is computed using normalized feature set. To define PDP, let  $S \subset \{x_1, x_2, \dots, x_p\}$  and  $C$  denote the complement set of  $S$  such that  $Z = S \cup C = \{x_1, x_2, \dots, x_p\}$  where,  $p$  is the total number of gait features. The term  $Z^S$  denotes the set of features for which partial dependence function needs to be plotted and  $Z^C$  demonstrates all the other features that were used in predicting the outcome of ML model (ensemble regression model). The ensemble regression (ER) model  $f(\cdot)$  is defined as combination of several decision trees to produce better predictive performance than utilizing a single decision tree. The various groups of learners are hereby combined to form a strong learner. Using bagging approach, several subsets of data from training dataset are created randomly, and each subset is used for model training. Thus, the average of all predictions from different decision trees is used being more robust than a single decision tree (hyper parameters for trained ER model include bagged approach with 30 as number of learners containing leaf size of 8 and 0.1 as learning rate). In this work, partial dependence represents relationship between features and response in a trained ensemble regression model; the single input (gait) feature of interest is taken at a time. Thus,  $S = \{x_1\}$  and  $C = \{x_2, x_3, \dots, x_p\}$  for computation of PDP for feature 1 (stance duration) and so on. Then,  $f(Z)$  provides partial dependency outcomes of trained ER model  $f(\cdot)$  to know the relationship between response (UPDRSm) and gait features that can be linear, monotonic, or complex. The subset of features considered is by marginalizing over other features in the set. Thus, a predicted outcome depends on all the variables in the set as

$$f(Z) = f(Z^S, Z^C) \tag{8}$$

To calculate PDP, selected subset consists of either one variable ( $Z^S$ ), while other variables are included in complementary set ( $Z^C$ ) in  $Z$  that are used in trained ER model. The partial dependencies of predicted outcome in  $Z^S$  are defined by expectation of the outcome with respect to  $Z^C$  using marginal probability. Each subset of features in  $S$  has its own partial dependence function  $f^S$ , which determines average value of  $f(Z)$  when  $Z^S$  is fixed, and other feature set  $Z^C$  varies its marginal distribution  $dP(Z^C)$ . Thus, by marginalizing over the other features, we get a function  $f^S$  that depends only on selected feature with inclusion of other feature interactions. The interaction between the two sets is weak, and all samples in the set are equally likely.



$$f^S(Z^S) = E[f(Z^S, Z^C)] = \int f(Z^S, Z^C) dP(Z^C) \tag{9}$$

The different values of  $Z_j^C$  are represented as  $\{Z_1^C, Z_2^C, \dots, Z_N^C\}$  occurring over all the observations in the training dataset. In other words, to estimate partial dependency of a considered feature (input feature of interest), the entire training set must be utilized for every set of calculated values. Finally, the partial function is the resultant of calculating averages in training data (Eqs. 8–10).

$$f^S(Z^S) \approx \frac{1}{N} \sum_{j=1}^N f(Z^S, Z_j^C) \tag{10}$$

This is a visualization tool, which assumes no or weak correlation of features in the given subsets  $(Z^S, Z^C)$ , and for one-dimensional feature set as in this case, PDP involves plotting of averaged value of  $f^S$  as a function of  $Z^S$  conventionally joined by a line.

Unlike PDP, which shows averaged relation between severity predication and gait features, the set of ICE plots shows individual dependency for each observation by disaggregating the averaged information. Thus, the contribution of each  $j^{th}$  sample is defined as follows from Eq. (11) by considering the estimated response function  $f^S$  and the observations. Considering each of  $N$  observed and fixed values of  $Z^C$ , a curve  $f_j^S$  is plotted against observed values in  $Z^S$ . Therefore, at each coordinate of x-axis,  $Z^S$  is fixed and  $Z^C$  varies across  $N$  observations in the training dataset. Therefore, each curve defines the conditional relationship between  $Z^S$  and  $f$  at fixed value of  $Z^C$ .

$$f_j^S = f(Z^S, Z_j^C) \tag{11}$$

In this work, average contribution of each extracted feature indicating the functional relationship between UPDRSm and normalized feature set is explored through PDP curves.

## 4 Results

### 4.1 Association between demography, anthropometry, walking speed, and gait features

The association of all spatio-temporal gait features (Table 5) with walking speed is high ( $|r_s| > 0.60$ ) except swing interval that shows a moderate correlation ( $|r_s| = 0.42$ ). However, after normalization through multi-variate model (highlighted in Section 3.3), all the parameter association reduced to less than  $|r_s| = 0.2$  (weak correlation). The gait features such as cadence, stride interval, swing interval, and step time before normalization are moderately related to gender, and their de-correlation significantly reduces association from  $0.17 < |r_s| < 0.34$  to  $0.002 < |r_s| < 0.06$ . The other measures such as age, height, and weight are shown to have moderate to weak correlation with gait features. Thus, after normalization significant correlations are successfully removed. All force-related features (Table 4) are normalized with weight of an individual and are not mentioned in Table 5.

All correlation between independent variables is computed using Spearman’s correlation except the dichotomous variable (gender) for which point bi-serial correlation coefficient [40] is used. The VIF analysis of each independent variable to identify dependencies among other individual independent measures is shown in Table 6.

Since all computed VIF values are less than 5, so all independent variables are incorporated into the regression model. The other diagnostic results to determine multicollinearity are included in Table 6 which comprises CI and VDP, although, VIF and CI indicate the presence of multicollinearity but cannot indicate which independent variables are multi-collinear. So, VDPs are used to identify such condition based on extent of inflation according to each CI.

The CI of more than 30 and VDPs of 0.9973 indicates linear dependencies in the regression model for one variable.

**Table 5** Spearman’s rank order correlation coefficient with pre (raw) and post normalization ( $p < 0.05$ )

Spearman’s correlations	Gender (G)		Age (A)		Height (H)		Weight (W)		Walking speed (S)	
	Raw	Norm	Raw	Norm	Raw	Norm	Raw	Norm	Raw	Norm
Gait features										
Cadence	0.31	0.018	0.07	0.089	-0.11	-0.063	-0.09	0.145	0.64	0.025
Stride length	-0.17	0.005	-0.15	-0.04	0.23	0.02	-0.03	-0.154	0.80	-0.003
Stride interval	-0.31	0.014	-0.08	-0.07	0.10	0.04	0.08	-0.151	-0.63	0.01
Swing interval	-0.33	0.004	-0.02	0.091	0.12	0.08	-0.08	-0.061	-0.42	0.037
Stance interval	-0.27	0.002	-0.07	-0.13	0.08	0.07	0.16	-0.049	-0.72	-0.004
Double limb support	-0.16	-0.05	-0.09	-0.19	0.04	0.011	0.30	0.069	-0.70	-0.186
Step length	-0.18	0.006	-0.16	-0.041	0.24	0.028	-0.04	-0.16	0.81	-0.004
Step time	-0.32	0.03	-0.09	-0.055	0.11	0.015	0.09	-0.165	-0.64	-0.013

The fact is also supported by moderate Spearman correlation between height with weight ( $R=0.55$ ), and with gender ( $R=-0.672$ ) (Table 6). Therefore, best-fitted regression model has not incorporated the height (H) into regression analysis. The resulting multi-variate models are shown in Table 8. As clear from the results, that self-selected walking speed (S) and gender (G) remained significant for all considered features. Due to moderate correlation between height (H) and weight (W), the W parameter remains significant during swing and double limb support, although, for swing interval, model performed with low adjusted  $R^2$  ( $\leq 0.39$ ) as compared to other parameters, but backward elimination considered this parameter at significance  $p < 0.001$ . Thus, all models and remained independent variables are tested for statistical significance ( $p < 0.001$ ). To avoid model

overfitting, RMSE-CV using tenfold is also computed. On observation, only slightly higher values are reported for RMSE-CV as compared to RMSE. Since, most of the gait features are correlated with self-selected walking speed along with demography and anthropometry, so it is advisable to consider walking speed as well to de-correlate data, which in turn, helps to reduce data dispersion.

The multi-variate approach succeeded in reducing the data dispersion as depicted by coefficient of variation (Table 9). The good predictive models such as stride length, stance interval, double limb support, and step length help to reduce dispersion by 51 to 67%. The other models that are cadence, stride interval, and step time reduced dispersion from 41 to 49%. On the other hand, less predictive model (swing interval) reduces dispersion by just 11%.

**Table 6** Spearman's rank order correlation coefficient and VIF for independent variables

Spearman's correlations	Gender (G)	Age (A)	Height (H)	Weight (W)	Walking speed (S)
Age (A)	0.098	1	-	-	-
Height (H)	-0.672	0.06	1	-	-
Weight (W)	-0.595	-0.01	0.55	1	-
Walking speed (S)	0.065	-0.14	0.07	-0.113	1
VIF analysis					
VIF	2.09	1.06	2.13	1.79	1.08

**Table 7** Multi-collinearity diagnostics

Multicollinearity diagnostics	Eigenvalue	Condition index (CI)	Variance decomposition proportions (VDPs)				
			Gender (G)	Age (A)	Height (H)	Weight (W)	Walking speed (S)
	2.2024	1	0.0033	0.0007	0.0002	0.0006	0.0006
	0.3354	6.5653	0.5524	0.0013	0.0015	0.0286	0.0009
	0.1384	15.909	0.0189	0.3065	0.0007	0.021	0.4522
	0.1189	18.5143	0.4165	0.4077	0.0003	0.5169	0.0944
	0.0566	38.896*	0.0089	0.2839	0.9973*	0.4329	0.4519

**Table 8** Multi-variate regression analysis for spatio-temporal parameters

Gait features (dependent variables)	Multi-variate regression model (independent variables)			Adjusted $R^2$	RMSE	RMSE-CV	
Spatio-temporal parameters							
Cadence	= 53.83	+ 36.27 S	+ 4.856 G	0.69	6.30	6.88	
Stride length	= 0.616	+ 0.6421 S	- 0.0571 G	0.77	0.070	0.078	
Stride interval	= 1.5833	- 0.3518 S	- 0.0471 G	0.61	0.06	0.08	
Swing interval	= 0.6283	- 0.0939 S	- 0.0333 G	- 0.001 W	0.39	0.02	0.03
Stance interval	= 1.045	- 0.2630 S	- 0.0282 G		0.71	0.038	0.04
Double limb support	= 0.2505	- 0.0921 S	- 0.0005 G	+ 0.0004 W	0.70	0.014	0.017
Step length	= 0.308	+ 0.3210 S	- 0.0286 G		0.76	0.036	0.039
Step time	= 0.791	- 0.1759 S	- 0.0236 G		0.66	0.03	0.04

**Table 9** Coefficient of variation (%) indicating data dispersion pre (raw) and post normalization

	Coefficient of variation (%)			Un-normalized data (raw data)			Normalized data			Significance ( <i>p</i> value)
	Mean	95% CL	SE	Mean	95% CL	SE	Mean	95% CL	SE	
Spatio-temporal parameters										
Cadence	9.29	[8.48:15.48]	1.31	4.99	[4.42:6.56]	0.69	4.99	[4.42:6.56]	0.69	<0.001
Stride length	17.35	[14.00:20.69]	1.69	5.76	[4.96:6.57]	0.61	5.76	[4.96:6.57]	0.61	<0.001
Stride interval	10.47	[8.64:12.29]	0.92	5.40	[5.09:7.71]	0.66	5.40	[5.09:7.71]	0.66	<0.001
Swing interval	11.02	[9.36:12.67]	0.83	9.78	[9.19:11.37]	0.39	9.78	[9.19:11.37]	0.39	<0.001
Stance interval	13.90	[12.33:15.47]	0.79	6.76	[5.21:7.31]	0.58	6.76	[5.21:7.31]	0.58	<0.001
Double limb support	29.80	[27.07:32.53]	1.42	13.7	[11.88:14.52]	0.36	13.7	[11.88:14.52]	0.36	<0.001
Step length	17.03	[15.68:19.02]	0.84	6.77	[5.97:7.58]	0.59	6.77	[5.97:7.58]	0.59	<0.001
Step time	24.24	[21.16:27.32]	1.55	12.28	[11.71:13.85]	0.79	12.28	[11.71:13.85]	0.79	<0.001

### 4.2 Assessment of Parkinson’s severity using normalized gait features

The proposed hybrid ML model provides good classification and assessment accuracies. Two experiments were carried out using two CV schemes. For tenfold CV prior to classification, the normalized dataset is randomly divided into two subsets: training data comprises 50 non-PD and 65 PD individuals containing 13,520 gait cycles from 2-min leveled ground walk and the testing dataset includes remaining 5800 gait cycles from another subject group not included as part of training (non-PD: 22, PD: 28). The tenfold CV is used for validation purposes in which the complete training dataset is randomly divided into ten equal parts. Further, each part (1352 gait cycles) is reserved for validation, and the remaining nine parts (12,168 gait cycles) are used for training by the considered ML model. The process continues until each part has been used for the validation. Thus, model training performance is the average score obtained in each tenfold. This approach extensively improves the model performance on unseen test data by removing bias and variance.

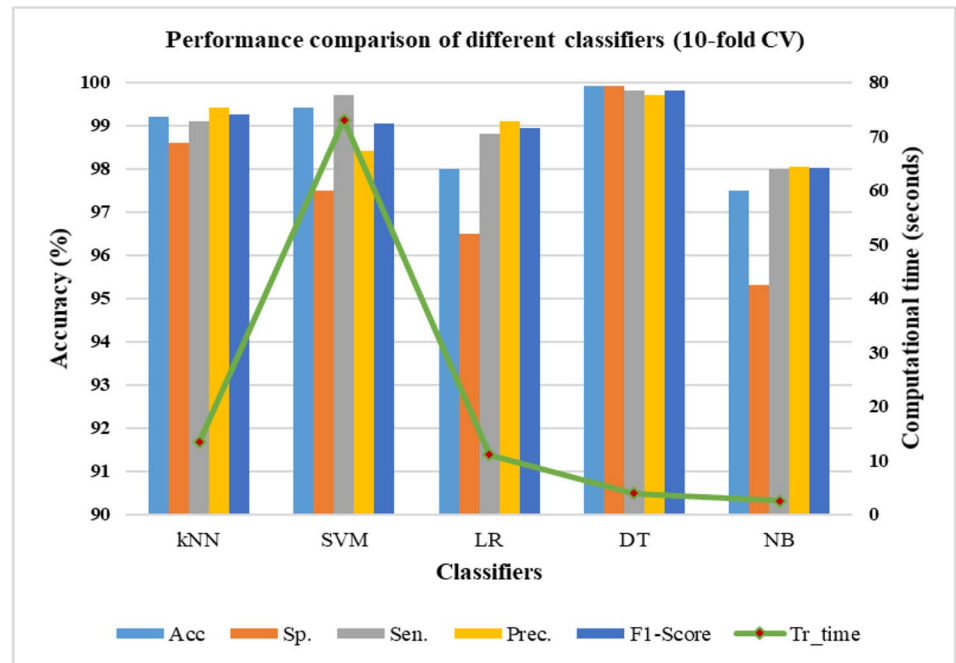
In leave-one-out CV, the validation process is based on leaving the gait feature dataset of one subject out for validation (as an unseen dataset) and using the gait features from other subjects for training the model. This process is repeated for total number of subjects in the dataset (165 subjects). The classification accuracy is the average of all the individually left-out subjects. The majority voting on prediction is done to make final prediction on subject because dataset contains on an average three walking trails for some subjects. The normalized features are fed to the hybrid model, and overall model evaluation is determined using metrics highlighted in Section 3.4.

The first stage of classification with tenfold CV comprising of DT classifies PD with mean accuracy of  $99.9 \pm 0.025\%$  using normalized feature set. Apart from acc., other metrics sp. and sen. performed equally well for DT classifier,

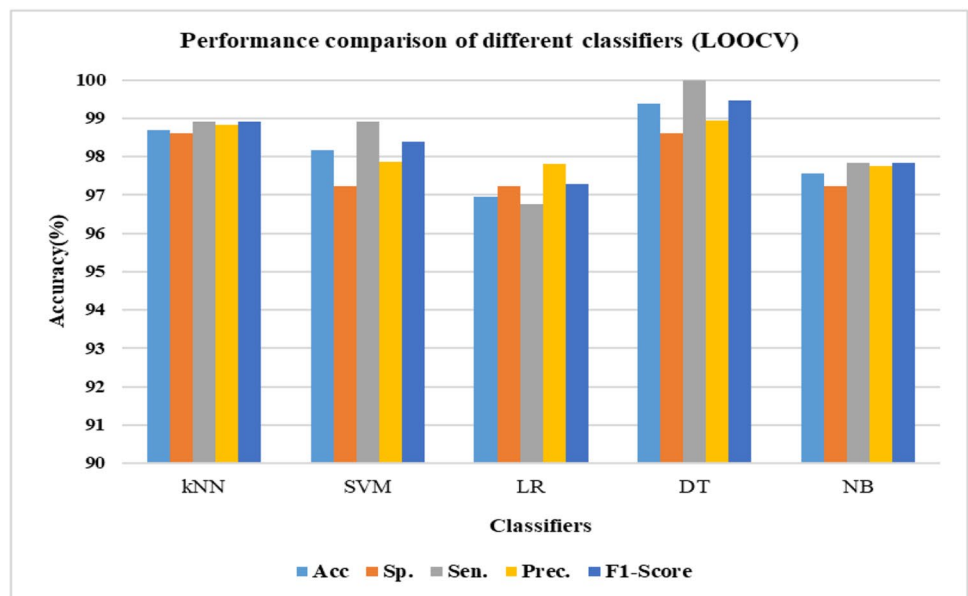
resulting in high positive predictive values and lesser false discovery rate for both the groups. Then, PD severity (UPDRSm) is determined using ER with mean RMSE of  $0.977 \pm 0.06$ , MAE =  $0.3476 \pm 0.024$ , and  $R^2$  as 98.7%. The results showed high inference accuracy with minimum misclassification rate as 0.3476 points (out of 108). It is also computationally efficient with average hybrid model training time (14.57 s) in comparison to deep-learning techniques. The PD and its severity assessment are also evaluated based on LOOCV using the same hybrid model which resulted in accuracy of 99.39% for PD prediction, and RMSE = 0.989, MAE = 0.3921, and  $R^2$  as 97% for PD severity prediction. It can be remarked from the results that proposed hybrid ML model well recognized the gait patterns. Nevertheless, our strategy requires domain knowledge for parameter extraction and normalization. Thus, this inherent knowledge in features suited to disease detection is valuable for model to achieve high inference accuracies. The classification results are computed using other classifiers such as SVM, kNN, LR, NB for comparison and analysis. The various performance metrics are evaluated and shown in Fig. 4.

The performance results using various evaluation metrics clearly demonstrate that DT classifier is robust performing consistently in terms of sp., sen., prec., and F1-score using both CV methods. However, SVM and kNN also presented good accuracies, but training time of these classifiers is higher as compared to DT for tenfold CV. The training time of NB is lower compared to DT, but the performance degrades for other evaluation metrics creating bias for detection accuracy of one class (non-PD). Thus, a trade-off between the two results in making DT as the possible choice for classification. All computations are executed on PC with Intel(R) Xeon(R) W-2155 CPU@ 3.30 GHz processor and 16 GB RAM. Thereafter, the classified PD individual is used as an input to second stage of hybrid model to determine PD severity comprising score from 0 (no disability) to 108

**Fig. 4** Comparative analysis of various classifiers for PD prediction. (a) tenfold CV. (b) LOOCV



(a)



(b)

(total disability). The regressor results are also computed using various regression approaches for comparison such as SVM regressor, Gaussian process regressor (GP), and DT regressor. From Fig. 5, it is clear that ER performed better in terms of  $R^2$  as well as reported low RMSE and MAE for both CV methods. Moreover, the training time of DT regressor is 3.52 s which is reduced by 69% in comparison with ER. But the RMSE and MAE reported

are higher as shown in Fig. 5. It can be demonstrated from the results that proposed model performed equally well in detection of PD and its severity using both CV approaches. But, as LOOCV is a special case of  $k$ -fold CV, where,  $k$  depends upon the total number of subjects in the dataset, it is computationally more expensive than tenfold but has led to creation of a generic model with more generalization ability as required for medical data.

**Fig. 5** Comparative analysis of various regressors for PD severity prediction. (a) tenfold CV. (b) LOOCV

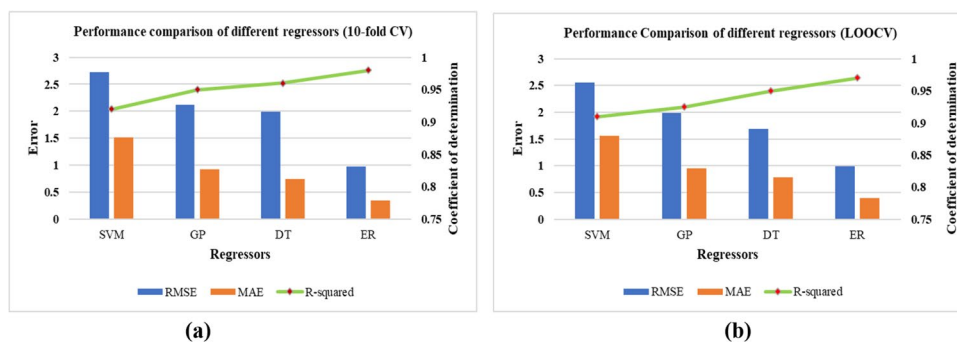


Figure 6 well demonstrated the capability of the model to determine PD severity for motor disorders.

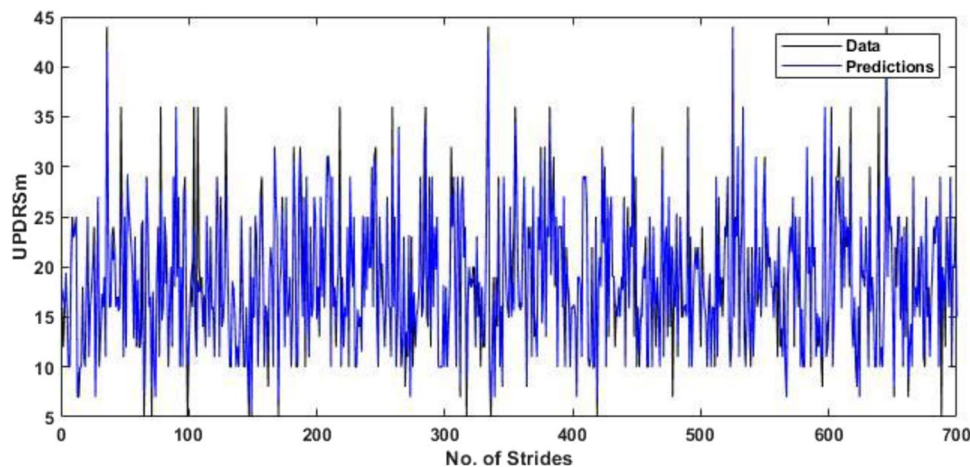
### 4.3 Interpretability of proposed hybrid strategy for clinical applicability

It has been observed when interpreting the gait features that consideration of multiple features is advantageous rather than focusing on single feature as it would be less accurate. Some features of PD gait when viewed individually overlap with non-PD individuals, particularly true in case of the initial stage of Parkinson’s. However, abnormalities can be detected if considered feature is accessed with other features altogether. As an individual tries to increase the pace, the average stride length and cadence increase. For instance, a shorter stride length with high cadence may be an attempt to achieve certain speed in which a faster rate compensates small steps that can be a sign of abnormality. Thus, these two features must be interpreted in reference to one other. A high cadence value combined with low stride length also indicates the difficulties in lifting the feet off the ground, which an individual compensates by taking faster steps. These

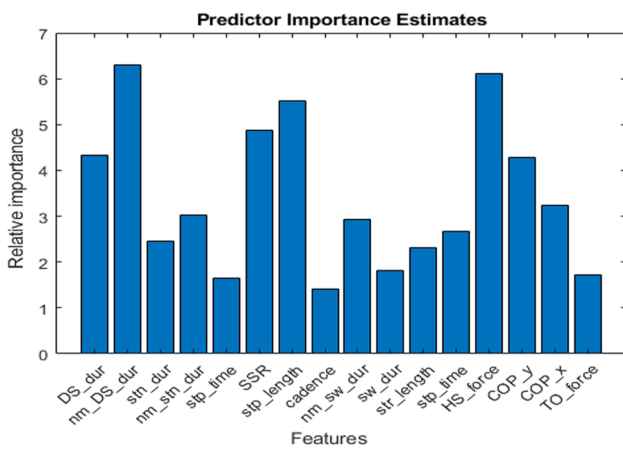
features can also be misleading to detect an abnormality if biased by individual age, gender, height, weight, and intended walking speed at that particular moment. Therefore, it becomes impossible to detect gait pathology by merely computing these gait features. Thus, it is always better to access the gait features with respect to subject’s physical properties. The significant inferences can be made by de-correlating the subjects’ data from the effect of anthropometry, demography, and self-selected walking speed of the individual. The greatest strength of this study is data normalization of clinically relevant features which aids in developing important insight to detect exact value of PD severity rating, and such information is useful in medical diagnosis.

The highly significant features for PD detection are step length (lower values), force variations at heel strike (less force values), COP\_y (more variability), swing stance ratio (lower values), and normalized (Nm) double support phase (higher values) as reported through feature importance analysis (Fig. 7). The heterogeneity and dependence of extracted features are determined by creating PDP curves (Fig. 8) indicating the functional relationship between extracted features (x-axis) and PD

**Fig. 6** PD severity assessment using UPDRSm





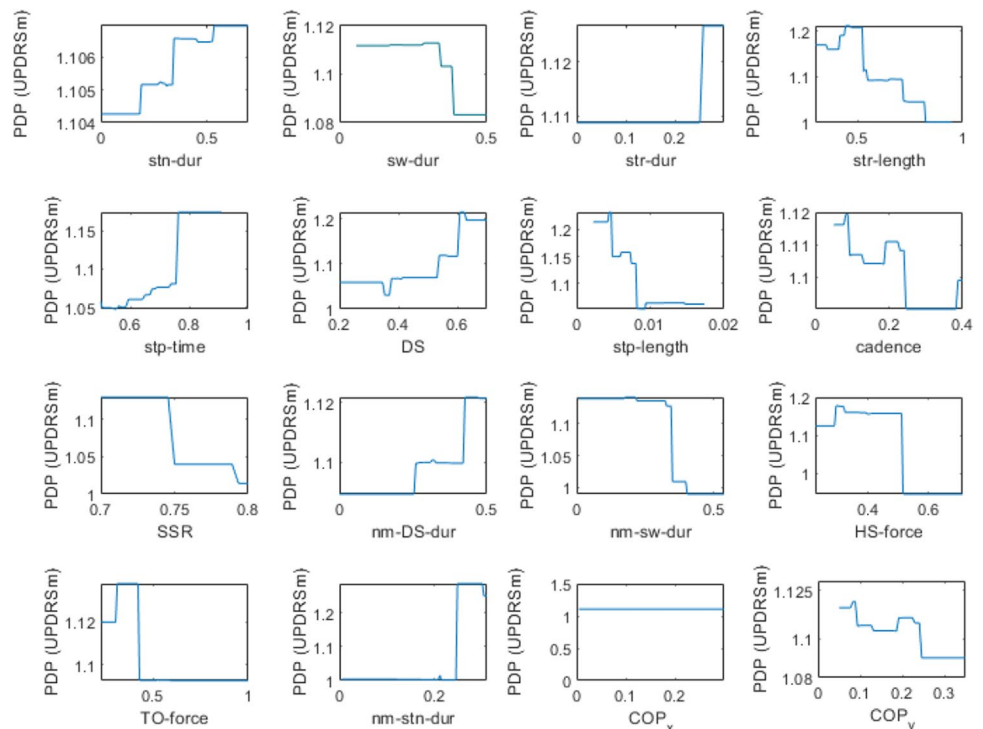


**Fig. 7** Feature importance estimates of extracted features

severity prediction (UPDRSm) (y-axis). These curves revealed that changes in normalized feature values do not exhibit a linear relationship with changes in severity prediction. The severity level UPDRSm (response) increases with stn-dur because it becomes harder for a PD patient to lift off the feet from ground; a significant variation is observed with an increase in severity thereafter. With high severity, reduction in sw-dur of PD individuals is seen, although a significant increase in str-dur is shown for high UPDRSm. The variability that is decrease in str-length and increase in stp-time with high severity is observed

from the trend line in Fig. 8. Nevertheless, a decrease is seen for step length with increase in severity which demonstrates more asymmetry with PD progression. The DS period is increased in case of high severity. Parkinson’s subject tends to avoid fall risk by reducing their sw-dur and increasing DS. But a small increase and then a decline are reported in number of steps for a given distance for high severity. A high cadence value combined with low str-length indicates difficulties in walking being compensated by faster steps. But sometimes when an individual is not able to take the steps faster that is in case of high severity index, then a decrease in cadence is observed. Yet a significant drop off with high severity level is observed for SSR indicating difficulty in initiation of walk for PD individuals and high asymmetry that alters this ratio. When the temporal features are normalized with respect to gait cycle/stride time (nm-DS-dur, nm-sw-dur, nm-stn-dur), the normalized values are very close to normal range that is, 0.6 over 1 for nm-stn-dur ( $0.6035 \pm 0.066$ ), and 0.4 over 1 for nm-sw-dur ( $0.395 \pm 0.037$ ) with the variability being very low. Any significant deviation from the nominal value indicates abnormality, though, higher deviation indicates severity and progression rate of the disease. This also accounts for poor coordination and loss of rhythmic movements leading to movement disorders. A reduction has been reported for force variation at time of HS as well as TO and shift in pressure profile (COP\_x, COP\_y) due to flat foot in PD subjects. In some cases, the heel strike

**Fig. 8** Partial dependency plots for extracted features



was not observed, and feet get strucked on the ground for longer duration (as depicted through high nm-stn-dur and low nm-sw-dur).

### 5 Discussion

The study is carried out to assess PD severity based on the proposed hybrid strategy using ML techniques and normalized feature set. The previously reported studies [15, 17] have shown considerable effects of individual’s demography, anthropometry, and walking speed on gait features. Using, dimensionless equations [31] cannot account for correcting multiple characteristics, thus, are proved to be less efficient. The multi-variate regression approach overcomes the said limitation by de-correlating these factors from gait features and improving the assessment. The correlation between various physical properties and walking speed was accessed using Spearman’s rank order correlation which described most of data dispersion in spatio-temporal features and prevented the gait impairment related differences between PD and non-PD from being determined. Moreover, the force-related features are mostly influenced by subjects’ body mass [18] and are normalized with body mass of individual in this study. Therefore, after normalization, it can be inferred from gait characteristics that PD individuals possess an increase in stride interval, which leads to an increase in the stance interval and reduction in swing interval. Such individuals also tend to increase their double limb support to avoid the risk of falls. A reduction in stride length, step length, and cadence is observed for PD group as compared to non-PD counterparts. A significant decrease in force at time of heel strike and toe off determines the flat foot profile of PD subjects. Also, the weight distribution is significantly different in both the groups. All these factors show variability with disease progression rate. The highly significant features for PD detection reported in

this work includes step length (lower values), force variations at heel strike (less force values), COP\_y (more variability), swing stance ratio (lower values), and Nm double support phase (higher values) through feature importance estimates. This may have implications in evaluation of disease-modifying interventions and early assessment of PD. Table 10 highlights the performance validation of our proposed method to state-of-art approaches using same gait dataset.

The PD classification accuracy observed in the current study is higher as compared to previous accuracies reported from the literature, though, many studies only considered detection of PD without quantifying level of impairment. In the present study, to avoid data under- and over-fitting and to get statistically unbiased results, two experiments were performed such as tenfold and LOOCV. The mean accuracy obtained are 99.39% and 99.9 ± 0.025% for LOOCV and tenfold respectively. Another study [20] demonstrated PD detection with an accuracy of 99%, Sen. = 97.8%, and Sp. = 99.5%, but following statistics (RMSE = 7.382, MAE = 4.462, and correlation coefficient as 0.895) are perceived for UPDRS severity rating without normalized feature set using locally weighted RF. The reason for high RMSE can be attributed to inclusion of non-motor symptoms as well that cannot be computed using instrumented analysis. Alternatively, one study reported classification with deep neural network using UPDRS and that too categorized in broad categories. This study reported [12] higher accuracy (98.5%) in terms of classifying a PD individual, but accuracy of 85.3% is observed for severity analysis. Nevertheless, the approach takes a long training time which leads to computational burden. On the other hand, this study provides low RMSE rate of 0.989 and 0.977, MAE = 0.3921 and 0.3476, and R<sup>2</sup> as 97% and 98.7% for LOOCV and tenfold respectively with high computational efficacy. It can be remarked from the results that proposed hybrid ML model well recognized the gait patterns. The partial dependence of extracted

**Table 10** Comparative analysis of reported results with previous studies

Study	Features	Classifiers	Prediction accuracy
Zhang et al., 2013 [25]	VGRF signals from heel and toe for both limbs	SVM, LC-KSVD, D-KSVD	81.53–83.44%
Ertugrul et al., 2016 [26]	Statistical features	BN, NB, MLP, PART, LR, RF, FT, and rule learner	87.58–88.88%
Alam et al., 2017 [29]	Time and frequency domain features	SVM, kNN, DT, RF	85.1–95.7%
Khoury et al., 2019 [28]	Time domain features	kNN, CART, RF, SVM, K-means, GMM	80–91%
Maachi et al., 2019 [12]	VGRF signals from foot sensors	1D-Convnets	85.3%, 98.7%
Khera et al., 2021 [13]	Time domain features	kNN, SVM, DT, RF	85–98.50%
Proposed method	Temporal, spatial, force features	DT, ER (LOOCV) DT, ER (tenfold)	99.39% 99.9 ± 0.025%

features is determined by creating PDP curves indicating the functional relationship between extracted features and PD severity prediction (UPDRSm). These curves revealed that changes in normalized feature values do not exhibit a linear relationship with changes in severity prediction.

Most of the subjects (74%) considered in this study are elderly ( $\geq 60$  years), and regression model built in geriatric population is not influenced by height and age. These models reported a low correlation of  $|r_s| < 0.25$  and have no significance with extracted gait features ( $p > 0.05$ ). For the considered population, only walking speed and gender remained as dominant features, but swing interval and double limb support seem to be significantly correlated with weight of an individual as well. As walking speed remains highly correlated ( $|r_s| \geq 0.80$ ) with stride and step length, PD individuals generally have low walking speed ( $1.033 \pm 0.205$ ) as compared to HC ( $1.24 \pm 0.159$ ). Thus, larger differences in walking speed have influenced these parameters. It was believed that de-correlating with such factors allows us to study significant variations between healthy and pathological gait. The other gait features shown in Table 5 are moderately correlated with walking speed, and consideration of these factors helps us to build clinical manifestations between the two groups. It has also been observed when interpreting the gait features, that consideration of multiple features is advantageous to detect abnormalities rather than focusing on single feature. When an individual attempts to increase the pace, the average stride length and cadence increase. For instance, a shorter stride length with high cadence may be an attempt to achieve certain speed in which faster rate compensates small steps, which can be a sign of abnormality. Thus, these two features must be interpreted in reference to one other. Moreover, with every 10-year increase in age the severity of disease is increased leading to more gait difficulty and worsening postural instability. The total UPDRS motor scale was higher in old-age PD than middle-age counterparts. The overall UPDRSm is higher in males by 15.2% than females. However, the mobility measures such as TUG are higher for pathological gait by 22.9% than normal gait which accounts for poor mobility and balance problems and are indicative of higher fall risk in PD individuals.

The main contributions of this study are as follows: Firstly, the primary reason of improved severity prediction is due to the data normalization and use of hybrid model that well recognizes the non-linearity between features and associated outcomes. Secondly, proposed hybrid strategy with instrumented gait analysis aided in determining exact value of PD severity rating based on motor disorders rather than obtaining the categorical results that define the severity levels. The precise prognostic

assessment together with clinically significant features can definitely assist clinicians in making decisions, planning interventions, and monitoring the progression of diseases.

Although, the proposed hybrid ML strategy with normalized feature set is able to achieve high inferences for PD severity prediction in comparison with other reported studies, some of the limitations and future perspectives are worth mentioning. First, the multi-variate approach includes only a specific group of subjects (mostly elderly). But, most of the individuals develop PD at elderly age. So, for this specific group the realized approach is appropriate. Second, the non-linear effects are not captured by this approach so transformation in another domain like exponential, quadratic, and polynomial can do better. Finally, the consideration of other factors like cognition, tremor signal, and muscle strength with advanced PD staging can play an important role in predicting gait differences.

## 6 Conclusion

Human gait pattern variability acts as a significant biomarker to diagnose PD as well as to determine progression of disease. Clinicians require physiological and neurological examination of the subject to determine the disease characteristics. This criterion is highly subjective and depends on expertise of the clinicians. Therefore, the proposed work envisaged to develop a hybrid strategy for determining PD severity rating using ML techniques and instrumented analysis. The publicly available database of VGRF signal is utilized for assessing the performance of the proposed hybrid strategy. Further, it can serve as an automated and assistive tool that can aid in PD diagnosis. The extracted clinically significant features are normalized using age, gender, weight, height, and self-selected walking speed of an individual. The multi-variate analysis approach is able to de-correlate the variation due to individuals' characteristics and self-selected walking speed. However, the approach is significant to account for pathology-related differences between PD and non-PD to be perceived. Moreover, the resulting normalized gait parameters augment the capability to compare individuals with distinct physical properties. The computationally efficient hybrid strategy used for disease classification is able to differentiate a PD individual with reported accuracy of 99.39% and 99.9%. Also, assessment of PD severity using UPDRSm had resulted in RMSE of 0.989 and 0.977 with 97% and 98.7% as  $R^2$  using LOOCV and ten-fold CV respectively. The primary reason for improved

severity prediction is attributed to consideration of interaction of multiple features for depicting outcomes rather than focusing on single feature and normalization of the dataset. The proposed hybrid model can benefit the clinicians with precise prognostic outcomes rather than binary diagnostic decision. However, rather than only determining the accuracy of the proposed model, feature importance and PDP curves are used to build significant insight for clinical needs. Thus, the study has important implications for investigation of VGRF data in PD detection and evaluation of its severity.

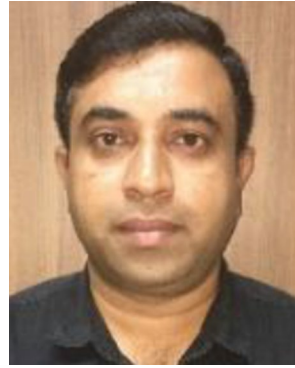
**Acknowledgements** The authors are thankful to Director, CSIR-CSIO, Chandigarh for extending the necessary support. Khera P. expresses gratitude to University Grants Commission, India for supporting her Ph.D. through its national fellowship program. The authors acknowledge PhysioNet for database support.

## References

- Hausdorff JM (2009) Gait dynamics in Parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 19(2): 026113
- Jankovic J (2008) Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry* 79(4):368–376
- Bhidayasiri R, Tarsy D (2012) Parkinson's disease: Hoehn and Yahr scale. *Movement disorders: a video atlas*. Springer, pp 4–5
- Goetz CG et al (2008) Movement disorder society-sponsored revision of the unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society* 23(15):2129–2170
- Rizzo G et al (2016) Accuracy of clinical diagnosis of Parkinson disease: a systematic review and meta-analysis. *Neurology* 86(6):566–576
- Hsu W-C et al (2018) Multiple-wearable-sensor-based gait classification and analysis in patients with neurological disorders. *Sensors* 18(10):3397
- Joshi S, et al. (2010) Classification of Alzheimer's disease and Parkinson's disease by using machine learning and neural network methods. in 2010 Second International Conference on Machine Learning and Computing. 2010. IEEE
- Amano S et al (2013) The effect of Tai Chi exercise on gait initiation and gait performance in persons with Parkinson's disease. *Parkinsonism Relat Disord* 19(11):955–960
- Pereira CR et al (2018) Handwritten dynamics assessment through convolutional neural networks: an application to Parkinson's disease identification. *Artif Intell Med* 87:67–77
- Sakar CO et al (2019) A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform. *Appl Soft Comput* 74:255–263
- Balaji E, Brindha D, Balakrishnan R (2020) Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease. *Applied Soft Computing*, 94: p. 106494
- El Maachi I, Bilodeau G-A, Bouachir W (2020) Deep 1D-convnet for accurate Parkinson disease detection and severity prediction from gait. *Expert Systems with Applications* 143: 113075
- Khera P, Kumar N (2021) Age-gender specific prediction model for Parkinson's severity assessment using gait biomarkers. *Engineering Science and Technology, an International Journal*
- Tropea TF, Chen-Plotkin AS (2018) Unlocking the mystery of biomarkers: a brief introduction, challenges and opportunities in Parkinson disease. *Parkinsonism Relat Disord* 46:S15–S18
- Wahid F et al (2016) A multiple regression approach to normalization of spatiotemporal gait features. *J Appl Biomech* 32(2):128–139
- Morris ME (2006) Locomotor training in people with Parkinson disease. *Phys Ther* 86(10):1426–1435
- Schwartz MH, Rozumalski A, Trost JP (2008) The effect of walking speed on the gait of typically developing children. *J Biomech* 41(8):1639–1650
- Mullineaux DR et al (2006) Normalization of ground reaction forces. *J Appl Biomech* 22(3):230–233
- Carty CP, Bennett MB (2009) The use of dimensionless scaling strategies in gait analysis. *Hum Mov Sci* 28(2):218–225
- Aşuroğlu T et al (2018) Parkinson's disease monitoring from gait analysis via foot-worn sensors. *Biocybernetics and Bio-medical Engineering* 38(3):760–772
- Frenkel-Toledo S et al (2005) Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society* 20(9):1109–1114
- Hausdorff JM et al (2007) Rhythmic auditory stimulation modulates gait variability in Parkinson's disease. *Eur J Neurosci* 26(8):2369–2375
- Yogev G et al (2005) Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding? *Eur J Neurosci* 22(5):1248–1256
- Perumal SV, Sankar R (2016) Gait and tremor assessment for patients with Parkinson's disease using wearable sensors. *Ict Express* 2(4):168–174
- Zhang Y, et al. (2013) Pathological gait detection of Parkinson's disease using sparse representation. in 2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE
- Ertuğrul ÖF et al (2016) Detection of Parkinson's disease by shifted one dimensional local binary patterns from gait. *Expert Syst Appl* 56:156–163
- Alkhatib R (2015) Gait-ground reaction force sensors selection based on ROC curve evaluation. *J Comput Commun* 3(03):13
- Khoury N et al (2019) Data-driven based approach to aid Parkinson's disease diagnosis. *Sensors* 19(2):242
- Alam MN, et al. (2017) Vertical ground reaction force marker for Parkinson's disease. *PLoS one*, 2017. 12(5): e0175951
- Zhao A et al (2018) A hybrid spatio-temporal model for detection and severity rating of Parkinson's Disease from gait data. *Neurocomputing* 315:1–8
- Wahid F et al (2015) Classification of Parkinson's disease gait using spatial-temporal gait features. *IEEE J Biomed Health Inform* 19(6):1794–1802
- Zeng W et al (2016) Parkinson's disease classification using gait analysis via deterministic learning. *Neurosci Lett* 633:268–278
- Prashanth R, Roy SD (2018) Novel and improved stage estimation in Parkinson's disease using clinical scales and machine learning. *Neurocomputing* 305:78–103

34. Goldberger AL, et al. (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 101(23): p. e215-e220
35. Sappakitkamjorn J, Niwitpong S-A (2013) Confidence intervals for the coefficients of variation with bounded parameters. *International Journal of Mathematical and Computational Sciences* 7(9):1416–1421
36. Rokach L, Maimon O (2005) Decision trees. *Data mining and knowledge discovery handbook*. Springer, pp 165–192
37. Elshawi R, Al-Mallah MH, Sakr S (2019) On the interpretability of machine learning-based model for predicting hypertension. *BMC Med Inform Decis Mak* 19(1):146
38. Molnar C (2020) Interpretable machine learning. Lulu. com. (PDP; Available as <https://christophm.github.io/interpretable-ml-book/pdp.html>)
39. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232
40. Brown, J.D., Point-biserial correlation coefficients. *Statistics*, 2001. 5(3).

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



**Dr. Neelesh Kumar** works as Sr. Principal Scientist in Department of Biomedical Applications of CSIR-CSIO, Chandigarh from 2001. He has completed ME (Instrumentation and Control) and PhD in gait analysis for prosthetic development. Presently he is working on development of exoskeleton devices for gait rehabilitation. His areas of interest are gait assessment, signal processing, artificial intelligence, design and development of assistive devices and methods to quantify rehabilitation.



**Preeti Khara** works as Senior Research Fellow (UGC-NET) in the Department of Biomedical Applications, CSIR-CSIO, Chandigarh, India. She has completed her graduation and post-graduation in the area of electronics and communication Engineering. Her research interest includes artificial intelligence, signal processing, gait analysis, machine learning, and gait rehabilitation.