Engineering Science and Technology, an International Journal 24 (2021) 112–125

HOSTED BY

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: www.elsevier.com/locate/jestch

Recognizing Parkinson's disease gait patterns by vibes algorithm and Hilbert-Huang transform

Fatih Aydın [a,*], Zafer Aslan [b]

[a] Department of Computer Engineering, Faculty of Engineering, Balikesir University, Balikesir 10010, Turkey
[b] Department of Computer Engineering, Faculty of Engineering, Istanbul Aydin University, Istanbul 34295, Turkey

ARTICLE INFO

Article history:
Received 30 March 2020
Revised 11 September 2020
Accepted 2 December 2020
Available online 13 January 2021

Keywords:
Machine Learning
Parkinson's Disease
Ensemble Learning
Hilbert-Huang Transform
Feature Engineering

ABSTRACT

Parkinson's disease (PD) is the second most common neurodegenerative disorder all over the world. There are resting tremor, bradykinesia, and rarely dystonia, all of which are motor symptoms, among the manifestations of PD. But the direct use of these motor symptoms for diagnosis can be misleading since PD can be confused with other Parkinsonisms and further disorders with a similar symptom. Therefore gait can be used, which has significant dynamics in the detection of PD and is an extremely complex motion. In this paper, we employed a state-of-the-art ensemble learning algorithm, called the vibes algorithm, and the Hilbert-Huang Transform (HHT) to recognize PD gait patterns. We extracted the features by the processing of the signals, which come from sixteen sensors on the bottom of both feet, through HHT and sixteen statistical functions. We then performed the two-stage feature selection process by using the vibes algorithm and the OneRAttributeEval algorithm. Finally, we exploited the vibes algorithm and the Classification and Regression Trees as a base learner to differentiate between patients with PD and the control group. The classification accuracy, sensitivity and specificity rates of the proposed method are 98.79%, 98.92%, and 98.61%, respectively. Moreover, we thoroughly contrasted our method with the previous sixteen works. The experiment results demonstrated that our method is high-performance and maintains stability. We also found out two unrevealed markers that could provide support in clinical diagnosis for PD apart from the classification task.

1. Introduction

Neurodegenerative diseases lead to the structure and function loss of the neurons, inclusive of their death. Parkinson's disease (PD) has the highest second prevalence in neurodegenerative diseases. PD was firstly described in the paper entitled "An Essay on the Shaking Palsy" by James Parkinson in 1817 [1]. The outstanding characteristic of PD is bradykinesia, which is defined as the deceleration of motions. In the early stage of the disease, the most apparent symptoms of bradykinesia are inabilities in gait, sitting, and speech [2]. PD patients are prone to fall. Therefore, gait disorders are one of the distinctive features in determining PD patients. Besides, the patients' quality of life is immutable while quantifying gait signals. Moreover, gait measurement provides conveniences to the patients while monitoring their state of health periodically, as well.

Physicians can decide a definitive diagnosis of PD through only a post-mortem examination [3]. PD is an insidiously onset and gradually progressive disease [4]. The symptoms of PD are categorized into two subgroups: motor and non-motor [5]. All symptoms are not able to coexist in each patient and the violence of them is not the same in each PD patient [4,6,7]. Thereby, we can say that the combination of symptoms of PD is different for each patient. The symptoms of PD typically begin with resting tremor or bradykinesia, and seldom with dystonia [2,8]. Hence, the diagnosis of PD can particularly be misleading in the initial stage. Clinicopathologic studies have shown that the symptoms of 10% – 24% of patients diagnosed as PD depend on further reasons [6,9]. Therefore, it is required to concentrate on the motor symptoms which are common for each PD patient.

We thoroughly reviewed all the other research works, which are the same scope as this study. Besides, we would like to indicate that the Parkinson datasets are the same, which were used in this study and the following studies. Lee and Lim [10] introduced a novel method to distinguish between idiopathic PD patients and healthy subjects by using Neural Networks with Weighted Fuzzy Membership (NEWFM) function. The method firstly derives new forty features from the gait characteristics of the subjects, by using the Wavelet Transform (WT) and various statistical methods (SMEs). Daliri [11] proposed a novel approach for separating PD

* Corresponding author.
E-mail addresses: fatih.aydin@balikesir.edu.tr (F. Aydın), zaferaslan@aydin.edu.tr (Z. Aslan).
Peer review under responsibility of Karabuk University.

https://doi.org/10.1016/j.jestch.2020.12.005
2215-0986/

patients from healthy individuals. This method extracts the mean frequency and the variance of frequency from the spectrum of the gait signals by applying the Short-Time Fourier transform (STFT) to them and then formed a single big histogram using the concatenation of these components. They selected the histogram bins by using Fisher's discriminant ratio later on. Finally, they computed the chi-square distance between the reduced histograms and employed a support vector machine for classification. Aydın and Aslan [12] introduced a new method finding out the patterns in the data by using the aggregating one-dependence estimators (A1DE) algorithm and the wavelet transform, whose wavelet type is daubechies3. Alafeef and Fraiwan [13] introduced a new method based on gait analysis, Continuous Wavelet Transform (CWT), and ANN to classify PD subjects and healthy control subjects. Besides, they have exhibited the gait patterns recognizing PD severities based on the Hoehn and Yahr scale. STFT and WT used in these studies assume that a signal is linear. However, most of the signals in the real-world are both non-linear and unstable. Besides, the models built in both studies are black-box models because of the structures of the selected machine learning algorithms. In other words, such models cannot help humans in discovering new knowledge in the decision-making process.

Wahid et al. [14] developed a novel method based on Multiple Regression Normalization Strategy (MRNS) and Random Forest (RF). They used MRNS, which explains subject age, height, body mass, gender, and self-selected walking speed to detect differences in spatiotemporal gait features. The features that they used for classification are stride length, cadence, stance time, and double support time. Perumal and Sankar [15] presented a new method based on spatiotemporal and kinetic parameters of the gait, and Linear Discriminant Analysis (LDA). The spatiotemporal parameters comprise of the step distance, stride time, stance, and swing phases. The kinetic parameters consist of heel force below toe force, toe force, and the normalized ones of these three parameters. They also developed a method that is built on the frequency-domain characteristics of the signals and LDA for the tremor assessment. Abdulhay et al. [16] proposed a novel approach to diagnose PD. They developed the pulse duration algorithm (PDA) to extract temporal features such as the stance and swing phases and stride time. They then employed Medium Gaussian Support Vector Machine (MG-SVM) for gait pattern classification between PD subjects and healthy subjects. They also developed a tremor analysis for detecting the severity of PD. The spatiotemporal parameters and the kinetic parameters are employed to detect gait and balance disorders in general. The spatiotemporal parameters and the kinetic parameters are used to detect gait and balance disorders in general. It should be used more sophisticated signal processing methods to determine the difference between Parkinson's disease-driven gait disorders and other gait disorders.

Ertuğrul et al. [17] proposed a novel method that is built on the Shifted One-Dimensional Local Binary Patterns (Shifted 1D-LBP) and Multi-Layer Perceptron (MLP). They also pinpoint local changes in the gait signal by the Shifted 1D-LBP method. Zeng [18] introduced a new approach using deterministic learning theory for a distinction between healthy subjects and PD subjects. They exploited localized Radial Basis Function (RBF) networks for the identification of dynamics of the non-linear system. Medeiros et al. [19] proposed a new method based on the Principal Component Analysis (PCA) to extract the features and the Euclidean Distance (ED) to classify the data. Wu et al. [20] developed a Support Vector Machines (SVM) model that employs the approximate entropy (ApEn), normalized symbolic entropy (NSE), and signal turns count (STC) parameters to differentiate between PD patients and healthy individuals. These parameters give the measurements of stride fluctuations in PD. Nandy [21] introduced a novel approach based on statistical methods and Bayes' classifier for differentiating between PD patients and healthy subjects. They first utilized Normalized Auto Correlation (NAC) to compute the degree of fluctuation in the gait of the individuals. They then employed four different normality testing methods to obtain more informative features. For feature selection, they used the Fisher Discriminant Ratio (FDR).

Shrivastava et al. [22] effectively presented a comparative analysis of various algorithms inspired by nature for selecting the best features and increasing the classification accuracy of the Artificial Neural Network (ANN) algorithm used for distinguishing healthy subjects from PD subjects. The experimental results showed that the Binary Bat Algorithm (BBA) has better performance than conventional methods such as the Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and the Modified Cuckoo Search (MCS).

Some studies have been done for PD diagnosis based on deep learning networks. Zhao et al. [23] developed a novel model that put together a Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), which is a deep learning model, distinguishing PD subjects from healthy subjects. They also tried to classify PD according to the severity levels. Zeng et al. [24] proposed a novel method based on Three-Dimensional Phase Space Reconstruction (3D-PSR), Empirical Mode Decomposition (EMD), and ANN for distinguishing between PD patients and healthy subjects. El Maachi et al. [25] proposed a new approach based on a 1D-Convolutional neural network (1D-CNN) to construct a Deep Neural Network (DNN) classifier. They extracted the features by 18 parallel 1D-CNN. They tested their algorithm both to distinguish between PD individuals and healthy subjects and to predict the severity of PD by using the Unified Parkinson's Disease Rating Scale (UPDRS).

Some studies have been performed for PD diagnosis based on the Hilbert-Huang Transform. In the study done by Rojas et al. [26], a new Computer-Aided system for diagnosing Parkinson's disease is proposed, which is based on the Empirical Mode Decomposition that decomposes any non-linear and non-stationary time series into a small number of oscillatory Intrinsic Mode Functions a monotonous Residuum. In the study done by Karan et al. [27], empirical mode decomposition based features are used to reveal the speech characteristics, and a new feature called Intrinsic Mode Function Cepstral Coefficients (IMFCC) is introduced to find out the patterns of Parkinsonian individuals' speeches.

Finally, Farashi [28] introduced a novel method for discriminating between individuals with PD and healthy individuals by using time, frequency, and time–frequency domains extracted from vertical ground reaction force (VGRF) data and incorporating a decision tree classifier.

The contributions of the proposed algorithm compared to the other works in literature are as follows:

Performing high-performance and maintaining stability in the experiment results.

Building a transparent model helping physicians in discovering new knowledge by using the vibes algorithm [29] and Classification and Regression Trees (CART) [30].

The heels are a significant marker and more predictive in comparison with the other parts of the foot-bottom in differentiating PD from control (CO) subjects.

The left foot signals give more information as compared to the right foot signals in terms of the classification task.

It is also easy to implement the proposed method in hardware because it is less costly computationally.

The rest of the paper is organized as follows. In Section 2, we introduce various methods used in the proposed algorithm. In Section 3, we describe the methodology and experimental setup. In Section 4, we present results and discussion. Finally, we conclude in Section 5.

2. Preliminaries

2.1. The vibes algorithm

Ensemble learning is a machine learning paradigm where multiple learners are trained to solve the same problem. The basic principle of the ensemble learning approach is that the ensemble decision created by appropriately combined individual predictors is better than that of any ensemble member. Many experimental and theoretical studies [31,32] also prove this. Every machine learning algorithm has certain limits and therefore makes mistakes. The purpose of ensemble learning is to control the strengths and weaknesses of algorithms while making the best decision [33].

The vibes algorithm is an ensemble learning algorithm developed by Aydın and Aslan [29]. The vibes algorithm is based on majority-voting in terms of the model combination method, optimized forward search in terms of the model selection method, and it is homogenous in terms of base learner selection. Base learners utilize two fundamental methods handling a feature set. These are decision fusion and classifier selection. In the decision fusion, each base learner uses all the information of a feature set. In the classifier selection, each base learner treats like a local expert and merely uses a small set of features. The vibes algorithm applies both decision fusion and classifier selection methods.

The vibes algorithm relies on two assumptions. The first is that features are independent of each other. The latter is that features are dependent on each other. Let $\mathcal{D} = \{\mathcal{X}, \}$ be a training set, which consists of the input space $\in \mathbb{R}^{\times}$, where m denotes the number of data and n denotes the number of the features and the class labels .

Accordingly, the first assumption runs as follows: each base learner $\mathcal{L}, \in \{1, \cdots, \}$ learns a single feature on the training set, in addition to the original training set. A temporary ensemble model $H : \mathcal{L} \cup \cdots \cup \mathcal{L}_j$ is then composed of the base learners. As to another assumption, the features are firstly arranged from the highest to the lowest according to their amount of the information in terms of classification. Either the ReliefF algorithm [34] or Shannon entropy [35] can be chosen for quantifying the amount of the information. After sorting the features of the original dataset, each base learner learns new datasets that are composed of the first one, the first two, the first three, and the first and nth features of the ordered original dataset, in order. A temporary ensemble model $H : \mathcal{L} \cup \cdots \cup \mathcal{L}_j$ is composed of base learners afterward. Finally, k base learners are selected within $(n + 1)$ base learners to constitute the most suitable ensemble by using the majority-voting method and the optimized forward search algorithm (OFSA). The pseudo-code of the vibes algorithm is shown in Algorithm 1.

Algorithm 1: The vibes algorithm

Input:
X denotes a sample space
y denotes class labels
L denotes a base learner
f denotes a parameter about if the features are dependent on or independent
r denotes a parameter about what feature scoring method is selected (the ReliefF algorithm or Shannon entropy)
Output:
H denotes a final model
(1) Rank and score the features according to the parameter *r*
(2) **if** (***f*** is '*independent*') **:**
(3) Generate a new dataset per feature ($\boldsymbol{X}_i$, i = 1, . . ., n)
(4) Create new models $\boldsymbol{h_i}$ from the new datasets and ***X*** by using ***L*** and ***y***
(5) **else**
(6) Generate new datasets by the forms of $X_1, X_1 \cup X_n$, $X_1 \cup X_2 \ldots$
(7) Create new models $\boldsymbol{h_i}$ from the new datasets by using ***L*** and ***y***
(8) Select the most suitable models using OFSA
(9) Create the final model ***H*** based on the majority-voting of the models
(10) **return *H***

Without applying the optimization to the vibes algorithm, its time complexity is $\Omega(n)$ and $O(n^2)$, respectively for the best and the worst cases. The space complexity of the algorithm is $\Omega\left(\left(\max_{n,m\in\mathbb{N}^+}\{n,m\}\right)\right)$ and $\mathrm{O}\left(\left(\max_{n,m\in\mathbb{N}^+}\{n,m\}\right)^2\right)$ respectively, for the best and the worst cases [29].

An implementation of the vibes algorithm is available at https://uk.mathworks.com/matlabcentral/fileexchange/69350-the-vibes-algorithm.

2.2. Classification and regression trees algorithm

Classification and regression trees (CART) are a machine learning algorithm developed by Breiman et al. [30]. CART constructs simple but powerful models by partitioning the data space recursively for making up a prediction model from data. It adopts a greedy (i.e., non-trackback) approach in which decision trees are created in the form of a top-down recursively split and manage. The process of creating the decision tree is shown in Algorithm 2.

Algorithm 2: The CART algorithm

Input:
D denotes a training set

Output:
n denotes the root node

(1) Compute the weighted impurity of node ***n***
(2) Estimate the probability that an observation is in node ***n*** using $P(T) = \sum_{i\in T} w_i$ where w_j is the weight of observation ***i***, and T is the set of all observation indices in node n
(3) Sort p_i in ascending order ($\boldsymbol{p_i}$ denotes all the split predictors, $i = 1, \cdots, t$)
(4) Determine the best way to split node n using p_i by maximizing the impurity gain overall splitting candidates
(5) Split the node using the best split in step 4 if the stopping criteria are not satisfied
(6) **return *n***

The CART algorithm uses the Gini index for quantifying the impurity of the attributes. Let D be a dataset. Accordingly, the impurity of D is measured by the Gini index. The Gini index is calculated by using the formula in Equation (1). Let D be split in the way of D_1 and D_2 by using the attribute A, the Gini index is calculated by using the formula in Equation (2). Also, the reduction in impurity to be formed by a binary split on a discrete or continuous value feature is calculated using the formula in Equation (3).

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2 \quad (1)$$

where p denotes the probability of a case in a node.

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (2)$$

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \quad (3)$$

2.3. Hilbert-Huang transform

Traditional data analysis methods assume that a signal is linear and stationary. Wavelet analysis and Wagner-Wille distribution [36,37] assume that a signal is non-stationary but linear. Additionally, there exist lots of non-stationary time series analysis methods [38,39]. But, these methods assume that the system is stationary and stable. Indeed signals are both non-linear and unstable in many real systems, which are natural or human-made. Analyzing non-linear and non-stationary signals disparately than their nature engenders many troubles. The necessary and sufficient condition to represent non-linear and non-stationary data is to have an adaptive base. A priori defined function cannot recline on a strong and flexible base. Building an adaptive and posterior defined method is the key to solve such a problem [40]. Hilbert-Huang Transform (HHT) developed by Huang et al. can cope with such troubles [41–43].

HHT consists of Empiric Mode Decomposition (EMD) and Hilbert Spectral Analysis (HSA). HHT is a proper method for non-linear and non-stationary data analysis, particularly time–frequency-energy demonstrations. Furthermore, HHT was only empirically tested and gained validity [40].

EMD is a data-driven, posterior-defined, intuitional and adaptable method. EMD assumes that the data consists of different intrinsic oscillation modes. An intrinsic mode function (IMF) represents a simple oscillation mode corresponding to a simple harmonic function, that is, expresses the signal as a simple harmonic component, instead of constant amplitude and frequency. IMFs have an amplitude and frequency varying as the function of time. The original signal is the sum of IMFs along with the last monotonic tendency [42,44]. The pseudo-code of the EMD algorithm is shown in Algorithm 3.

Algorithm 3: The EMD algorithm

Input:
x denotes the original signal

Output:
d denotes IMFs
r denotes the residual signal

(1) $n = 0$ and find all extrema (maxima, minima) of $\boldsymbol{x} = r_0$
(2) Compute the interpolation of the local extrema of r_n for obtaining the lower (e_{min}) and upper (e_{max}) envelopes
(3) Compute the mean envelope ($m = \frac{e_{min}+e_{max}}{2}$)
(4) Compute the candidate IMF ($d_{n+1} = r_n - m$)
(5) **if** d_{n+1} is an IMF
(7) **else**
(8) Take d_{n+1} as the input in step 2
(9) Repeat the process until r_n satisfies the stop criteria

The local nature of EMD can generate oscillations in disparate scales for a mode or oscillations in similar scales for different modes. This is an unwanted situation, and using similar scales per the mode is desired. This outcome of EMD gives rise to a problem called "mode mixing". To eliminate the above-mentioned lack and many lacks such as being non-resistance to noise, Colominas et al. developed a method called the improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (the improved CEEMDAN) by improving EMD [44]. The pseudo-code of the improved CEEMDAN algorithm is shown in Algorithm 4.

Algorithm 4: The improved CEEMDAN algorithm

Input:
x denotes the original signal

Output:
d denotes IMFs
r denotes the residual signal

(1) Compute the local means of the realizations $x^{(i)} = x + \beta_0 E_1(w^{(i)})$ by EMD to obtain the first residual signal $\boldsymbol{r}_1 = \langle M(x^{(i)})\rangle$
(2) Compute the first mode $d_1 = x - r_1$
(3) Compute the second residue as the average of local means of the realizations $r_1 + \beta_1 E_2(w^{(i)})$ and compute the second mode $d_2 = r_1 - r_2 = r_1 - \langle M(r_1 + \beta_1 E_2(w^{(i)}))\rangle$
(4) Compute the kth residue for $k = 3, \cdots, K$ $r_k = \langle M(r_{k-1} + \beta_{k-1} E_k(w^{(i)}))\rangle$
(5) Compute the kth mode, $d_k = r_{k-1} - r_k$
(6) Go to step 4 for next k

$\beta_k = \varepsilon_k \sigma(r_k)$ is used to obtain the desired signal-to-noise ratio between the added noise and the residue to which the noise is added. $M(\cdot)$ denotes the operator which produces the local mean of the signal that is applied to. $E_k(\cdot)$ denotes the operator which produces the kth mode obtained by EMD. $w^{(i)}$ denotes a realization of zero mean unit variance white noise.

In all implementations, we used the EMD toolbox available at http://perso.ens-lyon.fr/patrick.flandrin/emd.html. An implementation of the original formulation of CEEMDAN can be found at http://www.bioingenieria.edu.ar/grupos/ldnlys/.

HHT aims to find out the relation between the time–frequency-energy of a signal. The method also locally depicts non-stationary data. To calculate Instantaneous frequency (IF) and amplitudes, and to describe a signal locally, HHT employs Hilbert transform (HT). The Hilbert transform $y(t)$ of a signal $x(t)$ is calculated as in Equation (4).

$$\boldsymbol{y}(\boldsymbol{t}) = \frac{1}{\boldsymbol{\pi}} \boldsymbol{P} \int_{-\infty}^{\infty} \frac{\boldsymbol{x}(\boldsymbol{\tau})}{\boldsymbol{t} - \boldsymbol{\tau}} \boldsymbol{d\tau} \quad (4)$$

where P is the Cauchy principal value of the singular integral. As a result of this transformation, the phase angle of the signal $x(t)$ is shifted by ± 90° to expand the signal $x(t)$ to the complex plane. Then, with the combination of the signal $x(t)$ and the Hilbert transformation of $x(t)$, the analytical representation $z(t)$ of the signal $x(t)$ is obtained as shown in Equation (5).

$$\boldsymbol{z}(\boldsymbol{t}) = \boldsymbol{x}(\boldsymbol{t}) + \boldsymbol{i}\boldsymbol{y}(\boldsymbol{t}) = \boldsymbol{a}(\boldsymbol{t}) \boldsymbol{e}^{\boldsymbol{i}\boldsymbol{\theta}(\boldsymbol{t})} \quad (5)$$

where $i = \sqrt{-1}$. The instantaneous amplitude and instantaneous phase of the analytical function $z(t)$ are shown in Equation (6) and (7), respectively.

$$\boldsymbol{a}(\boldsymbol{t}) = \left(\boldsymbol{x}^2 + \boldsymbol{y}^2\right)^{1/2} \tag{6}$$

$$\boldsymbol{\theta}(\boldsymbol{t}) = \boldsymbol{tan}^{-1}\frac{\boldsymbol{y}}{\boldsymbol{x}} \tag{7}$$

As a result, the instantaneous frequency is calculated as shown in Equation (8). Finally, the instantaneous frequency is the change of the instantaneous phase of a signal over time.

$$\boldsymbol{\omega} = \frac{\boldsymbol{d\theta}}{\boldsymbol{dt}} \tag{8}$$

3. Methodology

3.1. The Parkinson dataset

We drew the Parkinson dataset (which consists of Ga, Ju, and Si datasets) [45–48] used in this study from the PhysioNET database [49]. The data consist of the vertical ground reaction force (VGRF) records for roundly two minutes on level ground. The subjects walked at their chosen pace. Thus, one can consider the force records as a function of time and location (e.g., stride time, swing time, and so.) [49]. This dataset includes gait measurements obtained from 93 idiopathic PD patients and 72 healthy subjects. There are eight sensors under the feet of each subject. The sensors measure the vertical ground reaction force (Newton unit) as a function of time. The output of the sensors has been recorded at 100 samples per second by digitizing. Moreover, the records involve two signals, which are the sum of the outputs of eight detectors. The descriptive information about this dataset is shown in Table 1. A sample vertical ground reaction force record for the Parkinson dataset obtained from the PhysioNET database is shown in Fig. 1.

Table 1
The descriptive information for the Parkinson dataset [49]

Subjects	Number of data	Age	Weight	Height	Gender	
					M	F
PD	93	66.30 ± 9.50	72.40 ± 11.96	167.44 ± 8.59	58	35
Control	72	63.68 ± 8.70	72.69 ± 12.42	168.25 ± 8.58	40	32

3.2. The creating and selecting of the features

Creating features by applying statistical analysis to unstructured or raw data is an important step in the machine learning field. Additionally, feature selection is a crucial sub-topic in the machine learning field. Feature selection is the task of selecting a subset of the features used in the model creation process. Feature selection is preferred due to three reasons: facilitating to interpret models easily [50], reducing the training time of classifiers, and avoiding errors resulting from variance [51].

There exist three general approaches for feature selection: the filter approach, the wrapper approach, and the embedded approach. The filter approach employs the general characteristics of the training data independently of machine learning algorithms [52]. The wrapper approach discovers the relation between relevance level and the best feature subset and searches an optimal feature subset adapted to a machine learning algorithm [53]. The last one of the approaches deals with that the classification algorithms perform good learning about which of the selected feature subsets. This approach resembles the wrapper approach, but, the learning process has an effect on the search process in an embed-

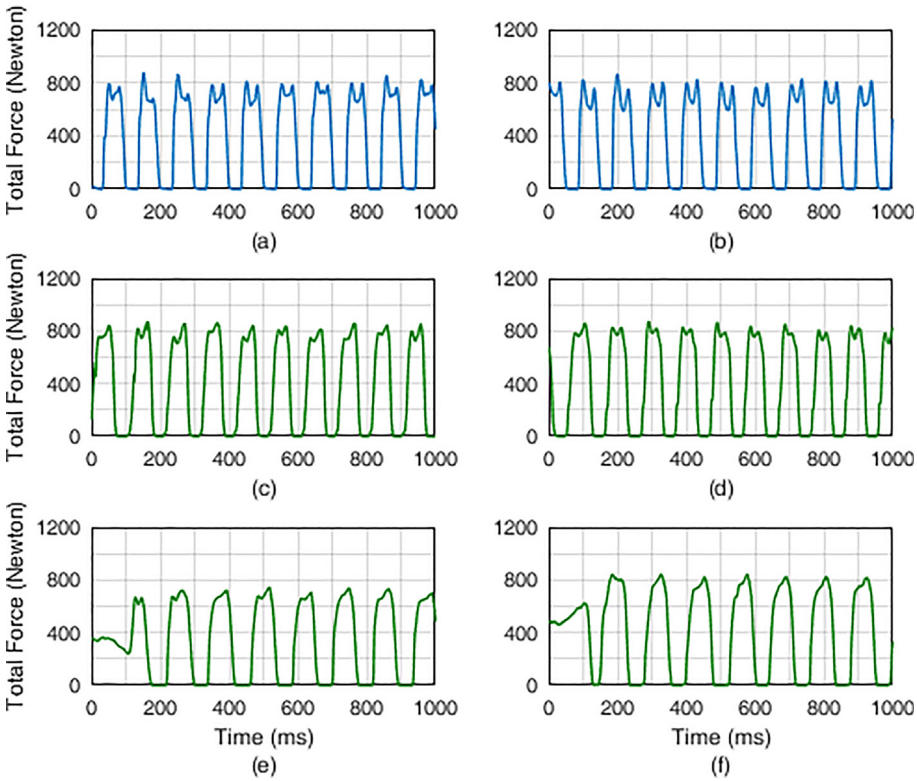

Fig. 1. The changes of the vertical ground reaction forces of the left and right feet of (a and b) a control individual, (c and d) a PD patient, and (e and f) other one PD patient over time.

ded approach. This situation reduces the computational cost and excessive trend.

There are many feature selection methods proposed in the literature. However, comparing these methods with each other is difficult. The reason is that datasets can contain many unrelated and redundant features. Thus, the performance of the feature selection algorithm is dependent on the performance of the learning algorithm. Additionally, there exist lots of performance criteria such as accuracy rate, computer sources, and the speed of feature selection in literature. Many researchers accept that there is no best method among feature selection algorithms [52]. Therefore, different feature selection algorithms try to address the problems that their domains involve by using different strategies.

In this study, we performed the feature creation and selection over the Parkinson dataset. We explained these steps to the following in detail. Each of the sixteen signals per the subject interacts with each other through three simple functions before beginning the feature creation. Let L1, L2, L3, L4, L5, L6, L7, L8, T1, R1, R2, R3, R4, R5, R6, R7, R8, and T2 be the eighteen signals obtained from each foot of the subjects. Two of these signals (i.e., T1 and T2) are the sum of the eight signals per foot. The 459 new signals are composed of the eighteen signals by using the three simple functions: $f_1(x,y) = x + y$, $f_2(x,y) = x - y$, and $f_3(x,y) = |x - y|$. For instance, we can obtain the three new signals, such as L1 + L2, L1-L2, and |L1-L2|, by applying the three functions to the signals L1 and L5. Consequently, we obtain the 477 signals per the subject after these operations.

We then obtain the first six IMFs of each signal by applying the improved CEEMDAN to the 477 signals. Moreover, we attain the amplitudes (A) and IFs of each IMF via HHT. The amplitudes, IFs, and IMFs belonging to the three subjects' left feet are shown in Fig. 2. We obtain Welch's PSD estimates of each signal later on. Thereby, we possess the IMFs, amplitudes, IFs, and Welch's PSD estimates per the signal. The Welch power spectral density estimates of the sum-signals belonging to the left and right feet of the three individuals are shown in Fig. 3. Besides, the feature creation\selection process is shown in Fig. 4.

We then apply the peak analysis to the amplitudes, IFs, Welch's PSD estimates, and the signals. We attain the four vectors corresponding to a signal by the peak analysis. These vectors are the maximum extreme values (pks), their locations (locs), their widths (w), and peak-height values (p). The peak analyses of the sum-signals belonging to the left and right feet of the three subjects are shown in Fig. 5.

Finally, we create the 694,512 features by applying the sixteen statistical functions to all the signals and vectors. These statistical functions are the coefficient of variation, the maximum value, the minimum value, the mean, the standard deviation, the coefficient of skewness, the coefficient of kurtosis, the median, the range, the Mean Absolute Deviation (MAD (0)), the Median Absolute Deviation (MAD (1)), the sum, the Shannon entropy, the log energy entropy, the sum of the absolute values of the vector elements, and the Euclidean length of the vector.

We proceed to the feature selection step after the feature creation step. We select the features by using the *OneRAttributeEval* method and run it over the stratified 10-fold cross-validation (CV) to ensure generalization performance. We then obtain the mean accuracy rates of each feature in the format of $(mean \pm deviation)$. Finally, we select the features corresponding to the rule specified by ourselves $(mean - deviation \geq 70)$. Thereby, we select only 135 features from 694,512 features.

3.3. The evaluation criteria

There need various statistical criteria to measure the performance of the classifiers at the end of the learning process. These criteria are the accuracy rate (ACC), the F-measure, the precision, the recall, the specificity (SPF), the Receiver Operating Characteristic (ROC) area, the Kappa value, the Root Mean Squared Error (RMSE).

The precision value indicates what percent of the instances predicted as positive are correctly predicted. The precision is also known as the positive predictive value (PPV). The recall value is also known as the sensitivity or the true-positive rate (TPR) and shows as positive are precisely predicted. The specificity value expresses what percentage of the instances labeled as negative are correctly predicted. The specificity is also known as the true-negative rate (TNR). The F-measure [54] is a score that considers both PPV and TPR of an experiment.

The erroneous predictions of a classifier result from either the random predictions or incorrect predictions [55]. As a result, RMSE can be used to quantify the performance of a classifier.

The Kappa value [56] is employed to measure the coherence between the predictions and actual values [57] and was first developed by Cohen. The Kappa value also gives information about how well a classifier learns the training data.

A classification of a new instance is obtained by comparing the score s of the instance with a 'classification threshold' t. If $s > t$,

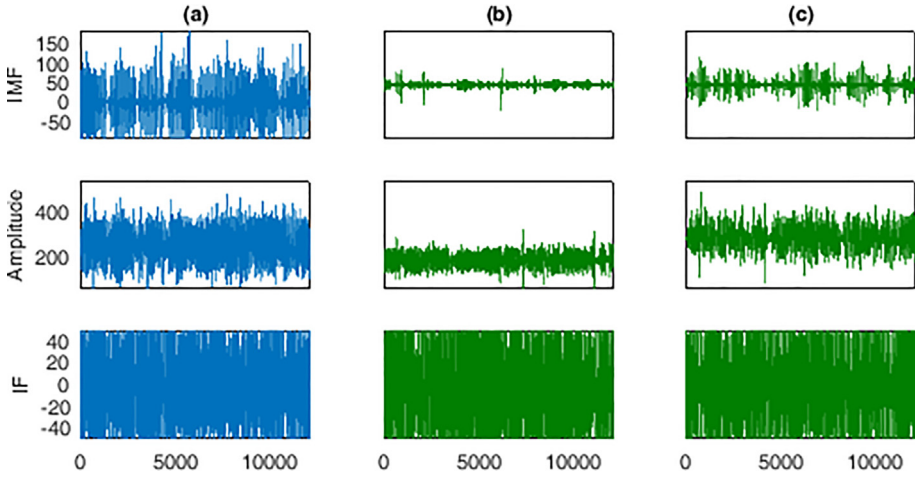

Fig. 2. The first IMFs of the vertical ground reaction forces of the left feet of (a) a control individual and (b and c) two PD patients, and the amplitudes and IFs corresponding to these IMFs.

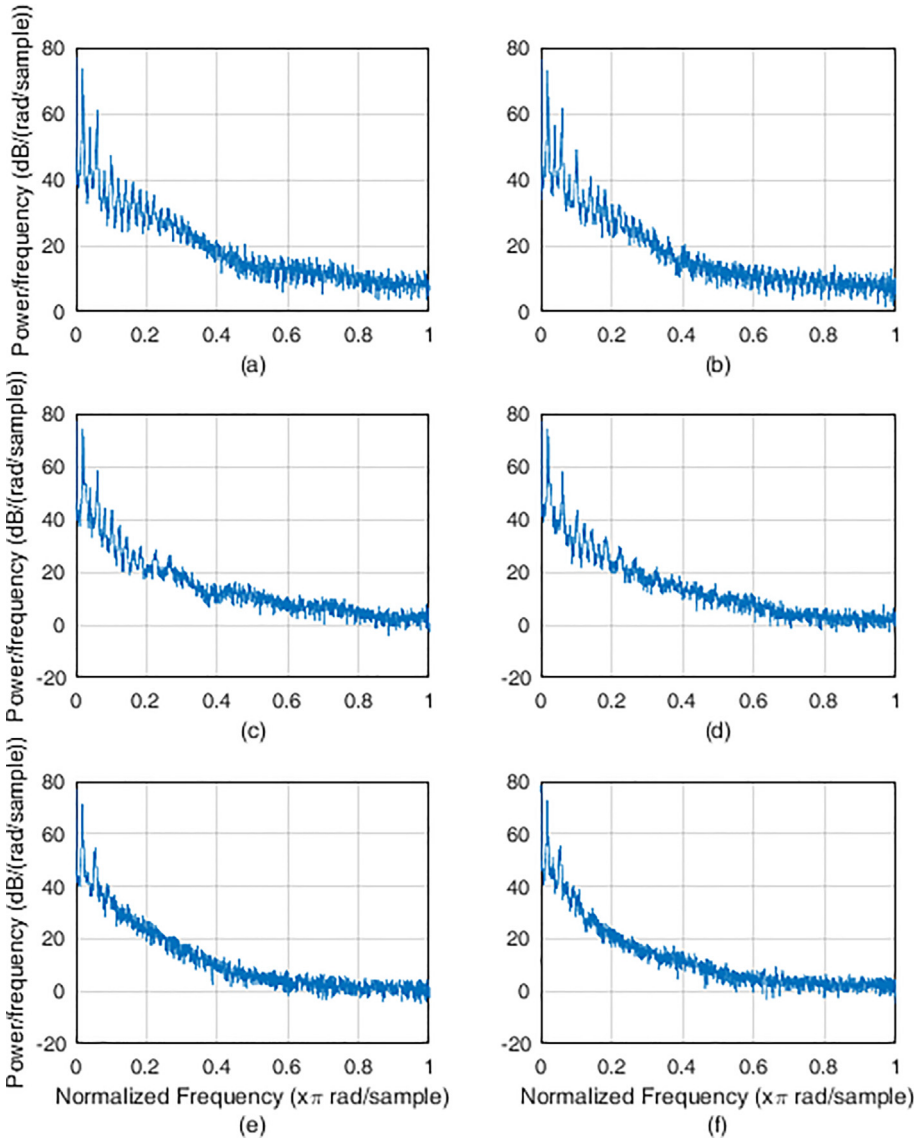

Fig. 3. The Welch power spectral density analyses of the vertical ground reaction forces of the left and right feet of (a and b) a control individual, (c and d) a PD patient, and (e and f) other one PD patient.

then the new instance is classified as coming from class 1, and if $s \leq t$ as coming from class 0 [58]. The ROC area is then the area under the ROC curve, which consists of a plot of the true-positive rate versus the false-positive rate (FPR) with the varying in the data amount.

4. Results and discussion

We performed all the experiments by employing the MATLAB R2016b tool and selected the CART algorithm as a base learner for the vibes algorithm. The changed parameters of the vibes and CART algorithms and their values are shown in Table 2. The rest parameters were initialized with default values. According to those parameter values, the vibes algorithm assumes that the features are independent of each other and uses the optimized forward search while searching the most suitable base learner models. Besides, the CART algorithm applies to the Twoing rule for deciding how to split a node.

We ran the vibes algorithm for recognizing the patterns in the Parkinson dataset, which has 135 features. We utilized 2-fold cross-validation (CV), 5-fold CV, 10-fold CV, and leave-one-out cross-validation (LOOCV) both to evaluate the performance of the vibes algorithm and to train it. By using the same dataset during both training and test, we also analyzed if the model suffers from underfitting or overfitting. Besides, each experiment, including training, was repeated 10-times. Thereby, the data are in a different fold in every repetition. The results of the experiments are shown in Table 3. Accordingly, we obtained the lowest and highest classification accuracy rates of 94.85% and 98.79% for 2-fold CV and LOOCV, respectively. Additionally, the mean classification accuracy rate is 96.2%, exclusive of all-training-all-test. We selected the model obtained in the LOOCV experiment to classify the PD and CO subjects because it has the highest ACC, TPR, and TNR, as well as without standard deviation (SD). The percentage of the PD subjects who are correctly identified is 98.92% for the LOOCV experiment. Furthermore, the percentage of the control subjects who are correctly identified is 98.61% for the LOOCV experiment. The model used 24 features of 135 features in the LOOCV experiment.

Feature selection algorithms are categorized into three types: the filter approach, wrapper approach, and embedded approach. The filter approach quantifies feature importance based on the statistical characteristics of the features. Therefore, it runs indepen-

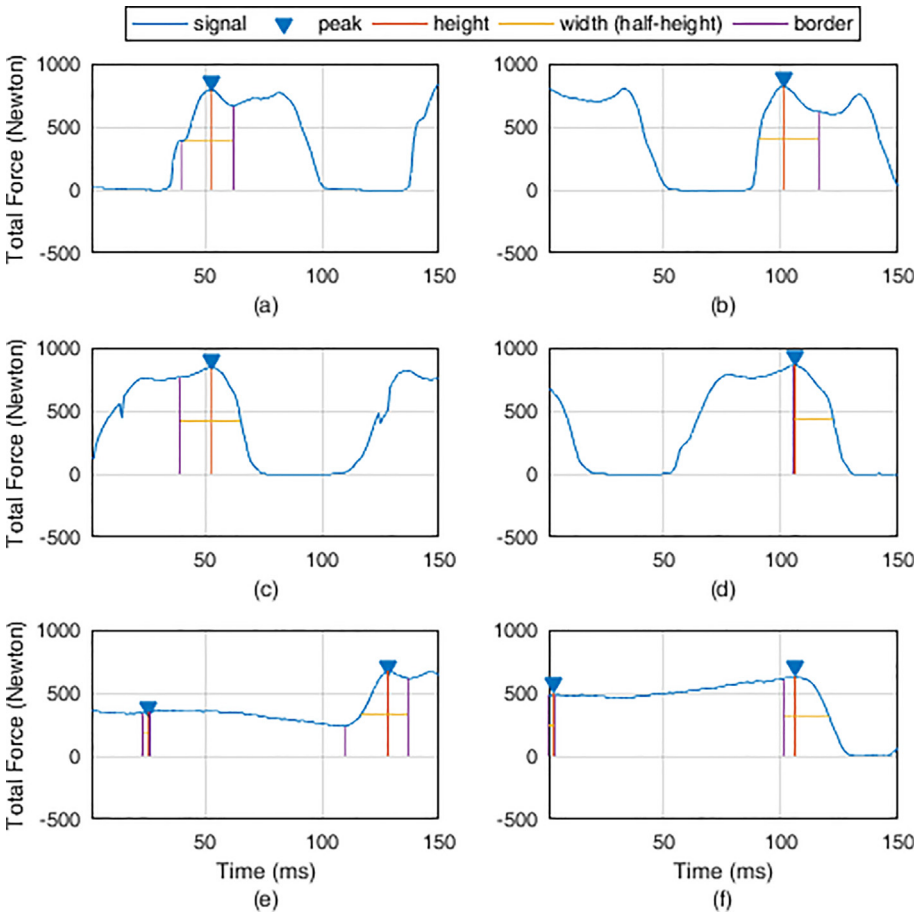

Fig. 4. The illustration of the feature creation and selection process.

dently of the training algorithm. To use the filter approach, we picked a function called Feature Selection using Neighborhood Component Analysis (fscnca) in the MATLAB toolbox. The function computes the feature weights by using a diagonal adaptation of neighborhood component analysis (NCA), i.e., measures feature importance by exploiting pairwise distances between data. The wrapper approach begins training by using a subset of features and then adds or removes a feature according to the selection criterion of the algorithm. It continues selecting the features until its stopping criteria are satisfied. To use the wrapper approach, we utilized a function called Sequential Feature Selection (sequentialfs) in the MATLAB toolbox, which builds a subset of features by sequentially selecting features until there is no improvement in the prediction of the selected classification algorithm. The embedded approach computes the scores of features as a part of the model learning process and selects features that work well together with the learning process. To use the embedded approach, we used a function called '*predictorImportance*' in the MATLAB toolbox, which computes predictor importance for the ensemble model by aggregating changes in the risk on each base learner in the ensemble. In terms of three feature selection types, the comparative results of various ensemble learning algorithms are shown in Table 4. According to these results, the embedded strategy puts together features making more accurate predictions on average on the data compared to the filter and wrapper approaches. Besides, the optimized forward search algorithm (OFSA) integrated into the vibes algorithm is better in terms of performance in comparison with the other feature selection methods. Exclusive of the Random Subspace learning algorithm, we have used CART as a base learner in all the ensemble learning algorithms used in the experiments, and the k-Nearest Neighbors algorithm as a base learner in the Random Subspace learning algorithm. Additionally, we have selected the method called '*twoing*' as the split criterion for CART, excluding the Adaptive Logistic Boosting and the Gentle Adaptive Boosting algorithms. For all the ensemble learning algorithms, the values of the rest parameters excluding '*twoing*' are default.

We plotted a learning curve shown in Fig. 6 (a) to give the average prediction quality of the LOOCV model according to the size of the training set. The prediction quality increases as the training set grow. Further, the average and last values of F-measure are 0.9769 and 0.9877, respectively after the number of data exceeds 12. Therefore, we can say that there are patterns in the data and the learning algorithm picks up them. The reason for the poor performance of a learning algorithm is either overfitting or underfitting the data. We plotted the curves of $RMSE_{LOOCV}$ and $RMSE_{training}$ shown in Fig. 6 (b) to estimate how well the LOOCV model matches the real function. If the origination of the error related to the LOOCV model is due to bias, the approaching of $RMSE_{LOOCV}$ and $RMSE_{training}$ to each other stops after a certain value as the amount of data is increasing. In other words, the bias error of the LOOCV model cannot be corrected even if the number of data is increased. Additionally, there is a slight difference between $RMSE_{LOOCV}$ and $RMSE_{training}$ at each point. If the origination of the error is due to variance, $RMSE_{LOOCV}$ and $RMSE_{training}$ start to approach each other as the number of data increases. On the other hand, the variance error does not always decrease as the size of the dataset increases. Accordingly, we can remark that the LOOCV model suffers from neither the bias error nor the variance error. Further, the average

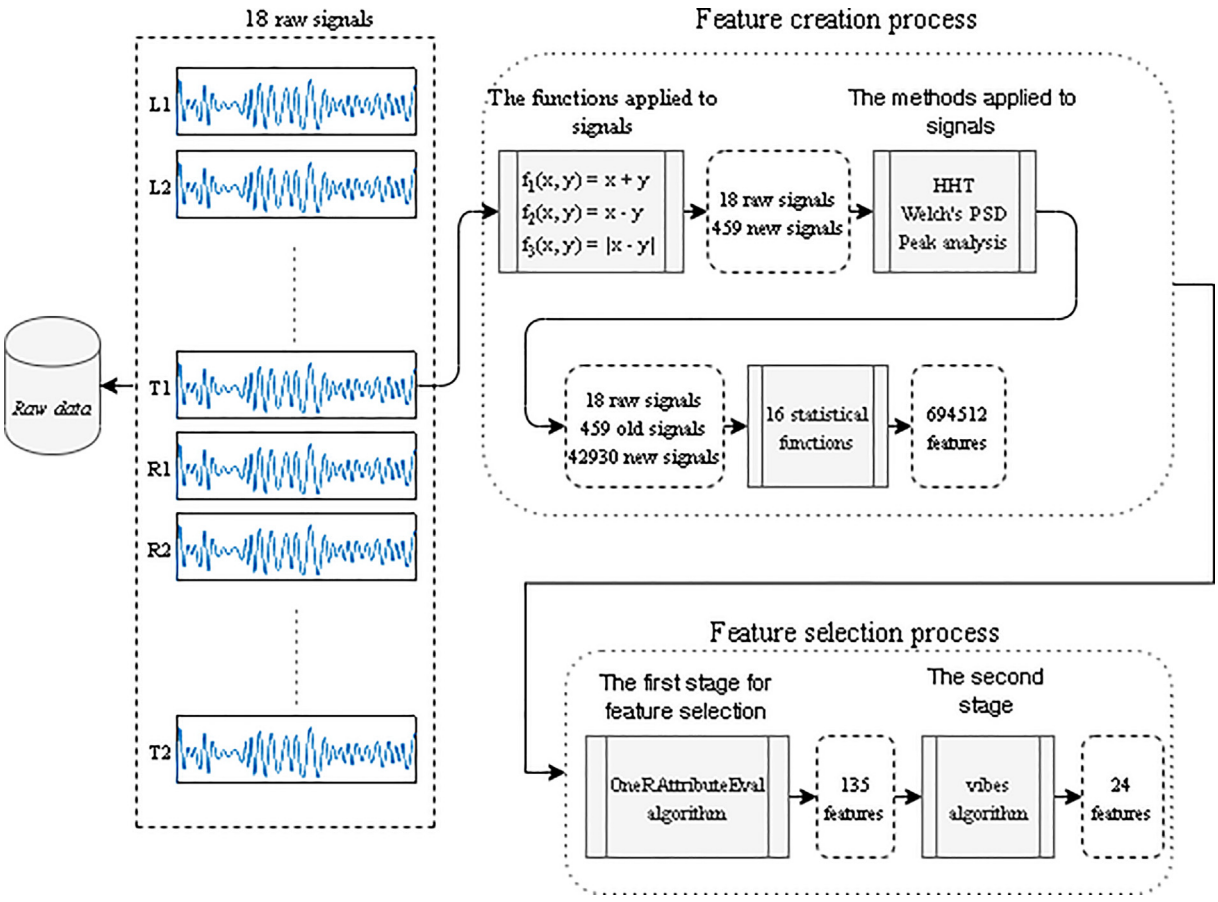

Fig. 5. The peak analyses of the vertical ground reaction forces of the left and right feet of (a and b) a control individual, (c and d) a PD patient, and (e and f) other one PD patient over time.

Table 2
The parameters and corresponding values of the vibes and CART algorithms.

Classifier	Parameter name	Value
Vibes	baseLearner	'tree'
	isDependent	false
	searchMethod	'OFS'
	featureRank	'IG'
CART	SplitCriterion	'twoing'

Table 3
The experimental results of the learning model.

Number of features	k-fold CV	ACC ± SD (%)	TPR ± SD (%)	TNR ± SD (%)
5 ± 0	all-training-all-test	100 ± 0.00	100 ± 0.00	100 ± 0.00
25 ± 9	2	94.85 ± 1.47	95.16 ± 2.32	94.80 ± 1.35
24 ± 8	5	95.52 ± 0.78	95.81 ± 1.70	95.47 ± 0.76
22 ± 6	10	95.64 ± 1.04	96.34 ± 2.05	95.53 ± 1.02
24 ± 0	LOOCV	98.79 ± 0.00	98.92 ± 0.00	98.61 ± 0.00

and last values of RMSEs of the LOOCV model are 0.1377 and 0.1101, respectively after the number of data exceeds 12.

The confusion matrix of this model is shown in Table 5. The model misclassified one PD subject and one CO subject. The data of these two subjects (i.e., 'SiCo17_01′ and 'SiPt35_01′) are obtained by Frenkel-Toledo et al. [45]. The other performance results of the model are shown in Table 6 and Table 7. According to those results, we can remark that the model well recognized the patterns in data because the Kappa value, the mean F-measure value, and the mean ROC area are very close to 1 and the RMSE value is close to 0. The model learns perfectly because the Kappa value is greater than 0.8. Moreover, the model can classify instances with both PD and CO very well because the average F-measure is very close to 1. The ROC area represents a measure of how well it can distinguish between PD and CO. Accordingly, the model can separate PD from CO because the average ROC value is very close to 1.

The selected features are shown in Table 8. The last column contains information about the classification accuracy rate of each feature itself. The operations in each column summarize how to generate a feature. For instance, we created the thirteenth feature as the following: we first obtained a new signal by subtracting the L3 signal from the L6 signal. We then acquired the fourth IMF of the new signal and extracted the IF signal from this IMF afterward. Finally, we achieved to form this feature by applying the entropy function with log energy to the peaks of this IF signal.

We have drawn the following statistical results from the features. The raw signals obtained from the subjects, by themselves, cannot help to recognize the patterns in the data. Considering the twenty-four raw signals used in the obtaining of the features, the left foot signals outnumber the right foot signals, (nearly 33.33%). Accordingly, we can say that the left foot signals provide more information in comparison to the right foot signals in terms of classification. Additionally, we most utilized the L1 signal while creating the features, in other words, it accounts for nearly 30% of all the signals. The R1 signal is also the second most used signal while creating the features, in other words, it accounts for nearly

Table 4
The comparative results of the ensemble learning algorithms with CART as a base learner, according to the feature selection types.

Classifiers	Embedded (OFSA)	Filter (fscnca)	Embedded (predictorImportance)	Wrapper (sequentialfs)
Vibes	95.64 ± 1.04	–	–	–
AdaBoostM1	–	85.86 ± 1.53	84.65 ± 1.85	82.42 ± 1.60
Bagging (Random Forest)	–	83.03 ± 1.05	81.62 ± 0.35	88.08 ± 0.35
Adaptive Logistic Boosting	–	88.28 ± 1.85	91.72 ± 0.93	79.39 ± 0.00
Gentle Adaptive Boosting	–	87.47 ± 3.11	90.30 ± 2.42	86.06 ± 2.10
Robust Boosting	–	83.23 ± 0.93	83.84 ± 1.85	81.41 ± 1.53
Random Undersampling Boosting	–	77.58 ± 2.19	81.62 ± 1.26	82.63 ± 2.29
Random Subspace	–	81.62 ± 0.70	–	84.24 ± 2.64

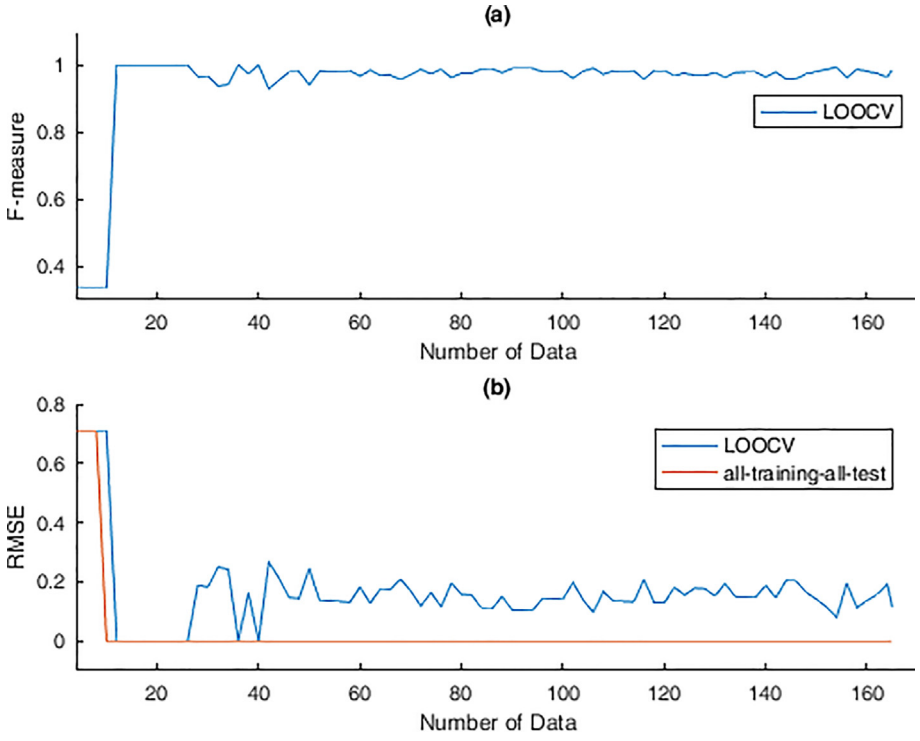

Fig. 6. (a) The learning curve of the LOOCV model (b) The variations of the RMSEs according to the number of data.

Table 5
The confusion matrix for LOOCV.

		Actual class	
		PD	CO
Predicted class	PD	92	1
	CO	1	71

Table 6
The basic statistics regarding the classification result for LOOCV.

Statistics	Value
The number of instances	165
Correctly classified instances	163
Incorrectly classified instances	2
Accuracy rate (%)	98.7879
RMSE	0.1101
Kappa	0.9754

15% of the used signals. Thus, the L1 and R1 signals account for nearly half of the signals creating the features. As a result, we would like to emphasize that the heels are a more important marker in comparison to the other parts of the foot-bottom in differentiating PD from CO subjects. In other words, the heel-strike and the heel-off stages of the gait cycle are much more predictive than the rest. We used Welch's PSD estimate in creating the seventeenth feature only. Therefore, we can state that Welch's PSD estimate does not contribute to recognizing the patterns as much as the improved CEEMDAN method. We employed the improved CEEMDAN method in 75% of the features. We particularly benefitted from the second IMFs of the signals. Considering the amplitude and IF signals extracted from IMFs, the amplitudes make the greatest contribution to classifying the data as compared to the IF signals. Furthermore, the maximum extreme values (pks) and peak-height values (p) are involved in the forming of the most suitable feature subset. No the locations (locs) and widths (w) of the peaks are involved in the forming of the feature subset. We especially took advantage of the max, the range, the mean, the median, the mean absolute deviation, and the median absolute deviation functions from sixteen statistical functions.

The comparative results of the designed method and other studies over the same Parkinson dataset are shown in Table 9. We scrutinized those within the below works, which are only close to our results. Furthermore, the criteria used in comparing our work with the previous works are accuracy rate, the true-positive

Table 7
The detailed accuracy table by the classes for LOOCV.

	TP	FP	PPV	TPR	TNR	F_1	ROC	Class
	0.9861	0.0108	0.9861	0.9861	0.9861	0.9861	0.9755	CO
	0.9892	0.0139	0.9892	0.9892	0.9892	0.9892	0.9755	PD
Weighted Average	0.9877	0.0123	0.9877	0.9877	0.9877	0.9877	0.9755	

Table 8
The features selected by the model for LOOCV.

No	Signal	Welch	IMF	A	IF	pks	locs	w	p	Function	ACC (%)
1	L1 - R2		2	✓						range	72.728 ± 1.310
2	L6 - R5									kurtosis	74.342 ± 2.264
3	L1 - L4		2		✓					median	71.920 ± 1.910
4	|R1 - R2|		2			✓				MAD(0)	72.526 ± 1.958
5	|L1 - L6|		1	✓					✓	median	72.732 ± 2.113
6	L1 + R1		2						✓	mean	72.525 ± 1.312
7	L1 + L2		3							MAD(1)	72.256 ± 2.117
8	L4 + L7									kurtosis	73.601 ± 1.060
9	|L1 - L3|		2						✓	sum	73.336 ± 2.630
10	R1 - R7									max	72.932 ± 2.391
11	L1 + R8		4							range	73.804 ± 2.230
12	L1 - R2		2	✓		✓				max	72.191 ± 2.033
13	L3 - L6		4		✓	✓				log energy	72.453 ± 2.215
14	L8 + R1		5	✓		✓				median	73.132 ± 2.674
15	L1 - L6					✓				mean	74.007 ± 1.970
16	|L1 - L3|		2	✓						Shannon entropy	73.340 ± 3.087
17	L1 - L6	✓								mean	72.997 ± 1.784
18	|L1 - L6|		2	✓					✓	MAD(0)	72.598 ± 2.525
19	|L1 - L3|		2						✓	Shannon entropy	72.933 ± 2.398
20	L8 + R1		5	✓						MAD(0)	75.015 ± 1.328
21	L4 + L8									skewness	73.739 ± 3.448
22	|R1 - R2|		2	✓						MAD(1)	73.329 ± 2.643
23	L1 - R4		1	✓		✓				MAD(0)	73.599 ± 2.035
24	|R1 - R6|		2	✓					✓	median	72.053 ± 1.950

rate, the true-negative rate, the stability in the experiment results, and the model complexity.

Accordingly, the model devised by [23] is more complex than ours. They used an LSTM with two-layers and set the number of the cell of each LSTM layer to 100 since the gait cycle of most subjects is about 1 s and the sampling frequency of their gait data is 10 ms. Each LSTM cell outputs a feature vector (which is 128-dimensional) for predicting the probability of its own softmax classifier. Their CNN channel consists of two convolutional layers, two max-pooling layers, and a fully connected layer. The convolutional layers output 32 and 64 feature maps, respectively. The fully connected layer outputs a 1024-dimensional feature vector to its own softmax classifier for PD detection. Finally, the weighted average of these two softmax classifiers was combined for the final classification result. The classification accuracy rate of this model is 98.61% under the 10-fold CV. We can remark that the implementation of this model to any hardware is difficult. We cannot say something certain about the real-world performance of the model because the TPR and TNR results of the model are unknown and the classification accuracy rates of the model were not carried out under the 2-fold CV, 5-fold CV, LOOCV, and all-training-all-test. They performed different training iterations in the three experiments to acquire the best result by setting up the iterations to 300000, 500000, and 500,000 respectively, but, the weighted average of the experiments should be taken for a more objective and certain result. Therefore, we can remark that this result may be a little biased. We ran each experiment for 10-times and gave the weighted average of these experiments together with their standard deviations. [25] stated a situation like this the following: there exists a discrepancy between the results obtained for the algorithm in [23] and the results reported in their paper.

Ref. [24] uses L5, L7, R7, and R8 signals because they have a significant VGRF difference between PD patients and healthy controls. They firstly computed the Euclidean distance of 3D phase space reconstruction of VGRFs of the L5, L7, R7, and R8 signals and concatenated these values to form a reference variable set. They then computed the EMDs of these reference variables and the derivation of gait features. They chose the third IMFs since most of the energy is concentrated in them. Finally, they classified the data by using an RBF Network. By using 2-fold, 10-fold, and leave-one-out cross-validation approaches, they obtained the classification accuracy rates of 91.46%, 96.99%, and 98.80%, respectively. The average classification accuracy rate, average TPR, and average TNR of these experiments is 95.75%, 95.66%, and 95.85%, respectively. Our results are 96.20%, 96.56%, and 96.10%, respectively. The highest classification accuracy rate, TPR, and TNR of them are 98.80%, 98.92%, and 98.63%, respectively. Our highest classification accuracy rate, TPR, and TNR are 98.79%, 98.92%, and 98.61%, respectively. Further, they chose the model giving the best result, but, the weighted average of all the experiments should be taken for a more objective and clear result. Therefore, we can say that this result may be a little biased.

Ref. [25] has used a DNN that consists of two parts. The first part consists of eighteen parallel 1D-CNN. The second part is a fully connected network that performs the concatenation of the extracted features. The final classification result was taken according to the majority-voting of all the segments. The Parkinson dataset also contains double-task signals, where subjects were walking while doing another activity. Accordingly, the number of data recorded in the Parkinson dataset is 306. They used just 300 of these data. We used the data of the subjects, which walk without doing another activity. Thus, we employed the 165 data in all. By

Table 9
The comparative results of the proposed method and other studies over the same Parkinson dataset.

Reference	Feature creation/extraction/ selection	Classifier	Number of features	Number of data	k-fold CV	ACC ± SD (%)	TPR ± SD (%)	TNR ± SD (%)
The proposed method	HHT	Vibes, CART	5 ± 0	165	all-training-all-test	100 ± 0.00	100 ± 0.00	100 ± 0.00
			25 ± 9		2	94.85 ± 1.47	95.16 ± 2.32	94.80 ± 1.35
			24 ± 8		5	95.52 ± 0.78	95.81 ± 1.70	95.47 ± 0.76
			22 ± 6		10	95.64 ± 1.04	96.34 ± 2.05	95.53 ± 1.02
			24 ± 0		LOOCV	98.79 ± 0.00	98.92 ± 0.00	98.61 ± 0.00
[28]	FFT, WT	A decision tree	27	165	10	100 ± 0.00	100 ± 0.00	100 ± 0.00
[25]	1D-CNN	1D-CNN	19	300	10	98.70 ± 2.3	98.10 ± 3.3	100 ± 0.0
[21]	FDR	Bayes	7	36	–	–	94.98	95.83
[13]	WT	ANN	–	166	training	99.10	–	–
					overall	97.60	–	–
					validation + testing	94.00	86.67	97.14
[24]	EMD, PSR	ANN	4	82	2	91.46	91.30	91.67
				166	10	96.99	96.77	97.26
				166	LOOCV	98.80	98.92	98.63
[12]	WT	A1DE	11	165	all-training-all-test	93.94	94.62	93.90
					2	90.91	91.40	90.90
					5	93.33	93.55	93.30
					10	93.94	94.62	93.90
					LOOCV	93.94	94.62	93.90
[23]	LSTM	CNN, Softmax	19	–	10	98.61 (all) 98.70 (Ga) 98.41 (Ju) 98.88 (Si)	–	–
[16]	PDA	MG-SVM	–	597	–	94.80	92.00	–
[20]	ApEn, NSE, STC	SVM	–	58	LOOCV	84.48	72.41	96.55
[22]	BBA	ANN	11	8000	all-training-all-test	99.00 (Ga)	99.30	98.60
				8922		98.30 (Ju)	98.50	98.20
				5500		98.50 (Si)	99.20	97.00
				8000	–	93.08 (Ga)	–	–
				8922		96.35 (Ju)	–	–
				5500		89.93 (Si)	–	–
[19]	PCA	ED	3	100	LOOCV	81.00	86.00	76.00
[15]	SMEs	LDA	10	40	5	92.50 (Ga) 92.50 (Ju) 90.00 (Si)	–	–
[18]	SMEs	RBF network	4	166	all-training-all-test	99.40	98.92	100
				83	2	91.57	91.49	91.67
				166	5	96.39	96.77	95.89
				166	LOOCV	93.37	92.47	94.52
[17]	Shifted 1D-LBP	MLP	108	308	10	88.89	88.90	82.20
[14]	MRNS	RF	4	49	–	92.60	80.00	70.00
[11]	STFT	SVM	21	166	2	91.20	91.71	89.92
[10]	WT	NEWFM function	40	1394	2	77.33	81.10	65.48

using the 10-fold CV approach, they obtained the classification accuracy rate of 98.70%. We obtained the classification accuracy rate of 95.94%. Furthermore, our highest classification accuracy rate is 98.79% by using the LOOCV style. The highest TPR and TNR of them are 98.10% and 100%, respectively. Our highest TPR and TNR are 98.92% and 98.61%, respectively. Accordingly, considering the standard deviations, our LOOCV model outperforms theirs in terms of the classification accuracy rate and TPR, i.e., the percentage of the PD subjects who are correctly identified. Finally, we would like to underline that the performance results (i.e., ACC ± SD and TPR ± SD are 98.70 ± 2.3 and 98.10 ± 3.3, respectively) that they obtained are statistically improbable.

Considering 2-fold CV, 5-fold CV, 10-fold CV, and LOOCV; the accuracy, sensitivity, and specificity rates of the proposed method are on average 96.20%, 96.56%, 96.10%, respectively. The results of the other works are the following: considering 10-fold CV; the corresponding rates of [25] are 98.70 ± 2.3, 98.10 ± 3.3, and 100 ± 0.0, respectively. Considering 2-fold CV, 10-fold CV, and LOOCV; the rates of [24] are 95.75%, 95.66%, and 95.85%, respectively. Additionally, there are no results regarding 2-fold CV, 5-fold CV, 10-fold CV, and LOOCV in most of the works. Moreover, there are no results either accuracy rate or sensitivity\specificity rates in some works. Hence, it is uncertain whether such works may satisfy the desired performance over real test data in the real-world.

Ref. [28] has performed a classification task by using the features obtaining from time, frequency, and time–frequency domains extracted from VGRF and the stance phase of the gait. The classification accuracies under 10-fold CV are 99.38 ± 1.98, 100 ± 0.0, and 100 ± 0.0, respectively, for the heel strike, the midfoot strike, and the forefoot strike of the left foot. The classification accuracies are 100 ± 0.0, 97.57 ± 3.13, and 100 ± 0.0, respectively, for the heel strike, the midfoot strike, and the forefoot strike of the right foot. However, some results are statistically unlikely because the sum of the standard deviations and accuracies of them is larger than 100. Besides, they did not report what kind of decision tree they used.

We maintained stability in the experiment results. The reason is that our results' standard deviation is 1.53 for 2-fold CV, 5-fold CV,

10-fold CV, and LOOCV. The standard deviation in [24] is 3.12. For [13], it is 1.8. In other words, the fluctuation in their standard deviations is large compared to ours. Furthermore, the standard deviation of our experiment is zero for LOOCV. That is, all the results are the same under 10 different cross-validation sets. Besides, the all-training-all-test results are not included in the results related to the standard deviation.

The proposed approach is computationally less costly than others because it was comprised of 24 decision tree models. The complexity of CART is $O(mn\log_2 n)$, where m denotes features and n denotes observations. Furthermore, the complexity of the vibes algorithm is $O(m^2)$ for the worst case. Neural network (NN) models in [13,24,25] have carried out lots of multiplications and the computational complexity of the total number of multiplications and additions is more than decision tree models.

5. Conclusion

A definitive diagnosis of Parkinson's disease (PD) is only determined by a post-mortem examination. PD is an insidiously onset and slowly progressive disease. PD has typical motor symptoms and non-motor symptoms. All of the symptoms do not coexist in each PD patient and the violence of the symptoms is not the same in each patient. Thus, it is noteworthy that the combination of symptoms and signs in PD, a heterogeneous clinical disease, is different in each patient. The symptomatic period of PD generally begins with resting tremor or bradykinesia, and more rarely with dystonia (involuntary contractions and seizures of the muscles). Because of all these reasons, the diagnosis of PD can be misleading, especially in the initial period. It has been demonstrated by clinico-pathologic studies that the symptoms of approximately 25% of patients who are pinpointed as PD depend on further reasons. The development of PD is initially mild and its symptoms may be different in each patient. Therefore, it is necessary to focus on the motor symptoms in which the symptoms are common. One of the parameters that can quantitatively measure the results of motor symptoms is kinetic (e.g., pressure and force). Gait is a dynamic that the important motor symptoms can uncover itself and enables us to quantify vertical ground reaction force (VGRF).

In this paper, we presented a state-of-the-art method to recognize the PD gait patterns. Our method relies on a novel ensemble learning algorithm (i.e., the vibes algorithm), Classification and Regression Trees (CART), and Hilbert-Huang Transform (HHT). Further, we rigorously performed feature creation and selection processes by exploiting HHT, the One-Rule Attribute Evaluation algorithm (OneRAttributeEval), and the sixteen statistical functions. The classification accuracy rate, true-positive rate (TPR), and true-negative rate (TNR) of the proposed model are 98.7879%, 98.92%, and 98.61%, respectively. Accordingly, our results are higher and more comprehensive as compared to the abovementioned studies, inclusive of some studies which have incoherent results. As a result of the features that we obtained, the heels are a significant marker and more predictive in comparison to the other parts of the foot-bottom in distinguishing between PD and control (CO) subjects. Besides, we revealed that the left foot signals give more information as compared to the right foot signals in terms of the classification task that and the L1 signal accounts for almost 30% of all the signals used the forming of the features. Additionally; attributes such as the amplitudes (A), the maximum extreme values (pks), and peak-height values (p) of the signals were overwhelmingly used as forming the selected features. The basic statistical functions such as the mean, the median, and the Mean\Median Absolute Deviation (MAD) were employed to form more than half of the selected features. It is likewise easy to implement the proposed technique in hardware because it is less costly, computationally. The model that we have obtained is composed of fusing 24 decision tree models. The decision trees result in a set of rules. These rules can be simply coded. Neural network (NN) models were created in [13,24,25] and many multiplications are performed in NN models. The run-time complexity of matrix multiplication is much more than the rule-based models.

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

[1] R.D. Adams, M. Victor, A.H. Ropper, Principles of Neurology, 6th ed., McGraw-Hill, New York, NY, 1997.
[2] F.B. Gibberd, The management of Parkinson's disease, The Practitioner. 230 (1986) 139–146. http://www.ncbi.nlm.nih.gov/pubmed/3703811.
[3] D.W. Dickson, H. Braak, J.E. Duda, C. Duyckaerts, T. Gasser, G.M. Halliday, J. Hardy, J.B. Leverenz, K. Del Tredici, Z.K. Wszolek, I. Litvan, Neuropathological assessment of Parkinson's disease: refining the diagnostic criteria, Lancet Neurol. 8 (2009) 1150–1157, https://doi.org/10.1016/S1474-4422(09)70238-8.
[4] J. Jankovic, A.H. Rajput, M.P. McDermott, D.P. Perl, The evolution of diagnosis in early Parkinson disease. Parkinson Study Group, Arch. Neurol. 57 (2000) 369–372, https://doi.org/10.1001/archneur.57.3.369.
[5] K.R. Chaudhuri, A.H. Schapira, Non-motor symptoms of Parkinson's disease: dopaminergic pathophysiology and treatment, Lancet Neurol. 8 (2009) 464–474, https://doi.org/10.1016/S1474-4422(09)70068-7.
[6] A.J. Hughes, S.E. Daniel, L. Kilford, A.J. Lees, Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases, J. Neurol. Neurosurg. Psych. 55 (1992) 181–184, https://doi.org/10.1136/jnnp.55.3.181.
[7] M. Lubarsky, J.L. Juncos, Progressive Supranuclear Palsy, Neurologist. 14 (2008) 79–88, https://doi.org/10.1097/NRL.0b013e31815cffc9.
[8] A. Deligtisch, B. Ford, H. Geyer, S.B. Bressman, Movement Disorders, in: J.C.M. Brust (Ed.), Current Diagnosis & Treatment Neurology, 2nd ed., Mc Graw Hill, New York, New York, USA, 2011, pp. 201–229.
[9] A.J. Hughes, S.E. Daniel, Y. Ben-Shlomo, A.J. Lees, The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service, Brain. 125 (2002) 861–870, https://doi.org/10.1093/brain/awf080.
[10] S.-H. Lee, J.S. Lim, Parkinson's disease classification using gait characteristics and wavelet-based feature extraction, Expert Syst. Appl. 39 (2012) 7338–7344, https://doi.org/10.1016/j.eswa.2012.01.084.
[11] M.R. Daliri, Chi-square distance kernel of the gaits for the diagnosis of Parkinson's disease, Biomed. Signal Process. Control 8 (2013) 66–70, https://doi.org/10.1016/j.bspc.2012.04.007.
[12] F. Aydın, Z. Aslan, Diagnosing Parkinson's Diseases by Machine Learning Methods Using Force Signals, in: Z. Aslan, F. Dökmen, E. Feoli, A.H. Siddiqi (Eds.), Mathematical Modeling of Real World Problems: Interdisciplinary Studies in Applied Mathematics, 1st ed., Nova Science Publishers, New York, 2019. https://novapublishers.com/shop/mathematical-modeling-of-real-world-problems-interdisciplinary-studies-in-applied-mathematics/.
[13] M. Alafeef, M. Fraiwan, On the diagnosis of idiopathic Parkinson's disease using continuous wavelet transform complex plot, J. Ambient Intell. Hum. Comput. 10 (2019) 2805–2815, https://doi.org/10.1007/s12652-018-1014-x.
[14] F. Wahid, R.K. Begg, C.J. Hass, S. Halgamuge, D.C. Ackland, Classification of Parkinson's Disease Gait Using Spatial-Temporal Gait Features, IEEE J. Biomed. Health. Inf. 19 (2015) 1794–1802, https://doi.org/10.1109/JBHI.2015.2450232.
[15] S.V. Perumal, R. Sankar, Gait and tremor assessment for patients with Parkinson's disease using wearable sensors, ICT Express. 2 (2016) 168–174, https://doi.org/10.1016/j.icte.2016.10.005.
[16] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan, V. Venkatraman, Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease, Future Gene. Comput. Syst. 83 (2018) 366–373, https://doi.org/10.1016/j.future.2018.02.009.
[17] Ö.F. Ertuğrul, Y. Kaya, R. Tekin, M.N. Almalı, Detection of Parkinson's disease by Shifted One Dimensional Local Binary Patterns from gait, Expert Syst. Appl. 56 (2016) 156–163, https://doi.org/10.1016/j.eswa.2016.03.018.
[18] W. Zeng, F. Liu, Q. Wang, Y. Wang, L. Ma, Y. Zhang, Parkinson's disease classification using gait analysis via deterministic learning, Neurosci. Lett. 633 (2016) 268–278, https://doi.org/10.1016/j.neulet.2016.09.043.
[19] L. Medeiros, H. Almeida, L. Dias, M. Perkusich, R. Fischer, A Gait Analysis Approach to Track Parkinson's Disease Evolution Using Principal Component Analysis, in: 2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2016: pp. 48–53. doi:10.1109/CBMS.2016.14.
[20] Y. Wu, P. Chen, X. Luo, M. Wu, L. Liao, S. Yang, R.M. Rangayyan, Measuring signal fluctuations in gait rhythm time series of patients with Parkinson's disease using entropy parameters, Biomed. Signal Process. Control 31 (2017) 265–271, https://doi.org/10.1016/j.bspc.2016.08.022.

[21] A. Nandy, Statistical methods for analysis of Parkinson's disease gait pattern and classification, Multimedia Tools Appl. 78 (2019) 19697–19734, https://doi.org/10.1007/s11042-019-7310-4.
[22] P. Shrivastava, A. Shukla, P. Vepakomma, N. Bhansali, K. Verma, A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease, Comput. Methods Prog. Biomed. 139 (2017) 171–179, https://doi.org/10.1016/j.cmpb.2016.07.029.
[23] A. Zhao, L. Qi, J. Li, J. Dong, H. Yu, A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data, Neurocomputing. 315 (2018) 1–8, https://doi.org/10.1016/j.neucom.2018.03.032.
[24] W. Zeng, C. Yuan, Q. Wang, F. Liu, Y. Wang, Classification of gait patterns between patients with Parkinson's disease and healthy controls using phase space reconstruction (PSR), empirical mode decomposition (EMD) and neural networks, Neural Net. 111 (2019) 64–76, https://doi.org/10.1016/j.neunet.2018.12.012.
[25] I. El Maachi, G.-A. Bilodeau, W. Bouachir, Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait, Expert Syst. Appl. 143 (2020), https://doi.org/10.1016/j.eswa.2019.113075 113075.
[26] A. Rojas, J.M. Górriz, J. Ramírez, I.A. Illán, F.J. Martínez-Murcia, A. Ortiz, M. Gómez Río, M. Moreno-Caballero, Application of Empirical Mode Decomposition (EMD) on DaTSCAN SPECT images to explore Parkinson Disease, Expert Syst. Appl. 40 (2013) 2756–2766, https://doi.org/10.1016/j.eswa.2012.11.017.
[27] B. Karan, S.S. Sahu, K. Mahto, Parkinson disease prediction using intrinsic mode function based features from speech signal, Biocybernetics Biomed. Eng. 40 (2020) 249–264, https://doi.org/10.1016/j.bbe.2019.05.005.
[28] S. Farashi, 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 (2020), https://doi.org/10.1016/j.bspc.2020.102132 102132.
[29] F. Aydın, Z. Aslan, The Construction of a Majority-Voting Ensemble Based on the Interrelation and Amount of Information of Features, Comput. J. 63 (11) (2020) 1756–1774, https://doi.org/10.1093/comjnl/bxz118.
[30] L. Breiman, J.H. Friedman, C.J. Stone, R.A. Olshen, Classification and Regression Trees, 1st ed.,., Chapman and Hall/CRC, 1984.
[31] T.G. Dietterich, Ensemble learning, in: M.A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, 2nd ed., MIT Press, 2002, pp. 405–408.
[32] C.-X. Zhang, J.-S. Zhang, A Survey of Selective Ensemble Learning Algorithms, Chin. J. Comput. 34 (2011) 1399–1410, https://doi.org/10.3724/SP.J.1016.2011.01399.
[33] Z.-H. Zhou, Ensemble Learning, in: Encyclopedia of Biometrics, Springer, US, Boston, MA, 2009, pp. 270–273, https://doi.org/10.1007/978-0-387-73003-5_293.
[34] I. Kononenko, Estimating attributes: Analysis and extensions of RELIEF, in: L. De Raed, F. Bergadano (Eds.), Machine Learning: ECML-94, Springer, Berlin, Heidelberg, Catania, Italy, 1994: pp. 171–182. doi:10.1007/3-540-57868-4_57.
[35] C.E. Shannon, A Mathematical Theory of Communication, Bell Syst. Tech. J. 27 (1948) 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.
[36] P. Flandrin, Time-Frequency/Time-Scale Analysis, Academic Press, 1999.
[37] K. Gröchenig, Foundations of Time-Frequency Analysis, Birkhäuser Boston, Boston, MA, 2001. doi:10.1007/978-1-4612-0003-1.
[38] H. Tong, Nonlinear Time Series Analysis Since 1990: Some Personal Reflections, Acta Mathematicae Applicatae Sinica, English Series. 18 (2002) 177–184. doi:10.1007/s102550200017.
[39] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, 2nd ed., Cambridge University Press, 2004.
[40] N.E. Huang, Introduction to the Hilbert-Huang Transform and Its Related Mathematical Problems, in: N.E. Huang, S.S.P. Shen (Eds.), Hilbert-Huang Transform and Its Applications, World Scientific Publishing, Singapore, 2005, pp. 1–26.
[41] N.E. Huang, S.R. Long, Z. Shen, The Mechanism for Frequency Downshift in Nonlinear Wave Evolution, Adv. Appl. Mech. 32 (1996) 59–117C, https://doi.org/10.1016/S0065-2156(08)70076-0.
[42] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceed Royal Soc. A Math. Phys. Eng. Sci. 454 (1998) 903–995, https://doi.org/10.1098/rspa.1998.0193.
[43] N.E. Huang, Z. Shen, S.R. Long, A New View of Nonlinear Water Waves: The Hilbert Spectrum, Annu. Rev. Fluid Mech. 31 (1999) 417–457, https://doi.org/10.1146/annurev.fluid.31.1.417.
[44] M.A. Colominas, G. Schlotthauer, M.E. Torres, Improved complete ensemble EMD: A suitable tool for biomedical signal processing, Biomed. Signal Process. Control 14 (2014) 19–29, https://doi.org/10.1016/j.bspc.2014.06.009.
[45] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, J.M. Hausdorff, Effect of gait speed on gait rhythmicity in Parkinson's disease: variability of stride time and swing time respond differently, J. Neuro Eng. Rehabil. 2 (2005) 23, https://doi.org/10.1186/1743-0003-2-23.
[46] G. Yogev, N. Giladi, C. Peretz, S. Springer, E.S. Simon, J.M. Hausdorff, Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding?, Eur J. Neurosci. 22 (2005) 1248–1256, https://doi.org/10.1111/j.1460-9568.2005.04298.x.
[47] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, J.M. Hausdorff, Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease, Mov. Disord. 20 (2005) 1109–1114, https://doi.org/10.1002/mds.20507.
[48] J.M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, N. Giladi, Rhythmic auditory stimulation modulates gait variability in Parkinson's disease, Eur. J. Neurosci. 26 (2007) 2369–2375, https://doi.org/10.1111/j.1460-9568.2007.05810.x.
[49] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J. E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals, Circulation 101 (2000) e215–e220, https://doi.org/10.1161/01.CIR.101.23.e215.
[50] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Springer, New York, New York, NY (2013), https://doi.org/10.1007/978-1-4614-7138-7.
[51] M.L. Bermingham, R. Pong-Wong, A. Spiliopoulou, C. Hayward, I. Rudan, H. Campbell, A.F. Wright, J.F. Wilson, F. Agakov, P. Navarro, C.S. Haley, Application of high-dimensional feature selection: evaluation for genomic prediction in man, Sci. Rep. 5 (2015) 10312, https://doi.org/10.1038/srep10312.
[52] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, A review of feature selection methods on synthetic data, Knowl. Inf. Syst. 34 (2013) 483–519, https://doi.org/10.1007/s10115-012-0487-8.
[53] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artif. Intell. 97 (1997) 273–324, https://doi.org/10.1016/S0004-3702(97)00043-X.
[54] C.J. Van Rijsbergen, Information Retrieval, 2nd ed., Butterworth-Heinemann, 1979.
[55] E.L. Lehmann, G. Casella, Theory of Point Estimation, Springer-Verlag, New York, 1998. doi:10.1007/b98854.
[56] J. Cohen, A Coefficient of Agreement for Nominal Scales, Educ. Psychol. Measur. 20 (1960) 37–46, https://doi.org/10.1177/001316446002000104.
[57] I.H. Witten, E. Frank, M.A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed., Morgan Kaufmann, 2011.
[58] D.J. Hand, Measuring classifier performance: a coherent alternative to the area under the ROC curve, Machine Learning. 77 (2009) 103–123, https://doi.org/10.1007/s10994-009-5119-5.