Empirical mode decomposition -Based Method for Parkinson Detection from Voice Signal

Sarab Jalal1, Sarmad K.D.Alkhafaji

College of Education for Pure Sciences, Computer science department, Thi-Qar University,

IRAQ*

Corresponding email: Dr.sarmad@utq.edu.iq, msc21co13@utq.edu.iq

Abstract:

Parkinson's disease (PD) is a brain disorder that causes speech and communication problems. The speech issues are often described as slow speech and difficulty with articulation because Jaw muscles don't move with enough strength. Mainly, clinical experts make a voice assessment to analyze voice signals, to detect PD. In this paper, we designed an intelligent model based on empirical mode decomposition (EMD) techniques for the detection of Parkinson's. Firstly, the voice signals are passed through EMD. The suggested EMD -based model is then used to extract a collection of entropy features from voice data. The features that have been chosen are fed into a K-nearest neighbor (KNN) as well as another four classifiers least squares support vector machine (LS-SVM), bagged tree, SVM (support vector machine), and Kmeans for the comparison. The proposed model is evaluated using a publicly available dataset named UCI machine learning repository. Several tests were carried out, and the findings revealed that the proposed framework can classify voice signals with a 97% accuracy in the K-nearest neighbor (KNN) technique.

Keywords: Parkinson's disease, EMD, KNN, entropy features.

1-Introduction

Parkinson's disease (PD) is a neurological disorder that weakens a patient's capacity to maintain normal motor function [1]. Around 7–10 million people globally are diagnosed with PD, including more than 3% over 75 and more than 1% over 55. According to WHO, the number of persons diagnosed with Parkinson's disease will double in 2030. It is crucial to detect Parkinson's disease as early as possible to slow its

development. With PD, the brain's nerve cells become impaired and they lose the ability to produce one of the important chemicals named dopamine. A low level of dopamine can cause movement-related problems, such as rigidity, tremors, and poor balance. Experts have not identified the precise cause of PD. However, many present studies showed that PD can result from a combination of genetics and environmental factors.

Currently, there is no specific assessment to diagnose PD. Experts may diagnose PD using a neurological exam, medical history, blood, voice, and brain assessments. Clinical reports showed that most patients with PD showed voice disorders by the time they diagnose with PD symptoms.

In recent years many researchers have explored the correlation between vocal changes and Parkinson's. Compared with other patients' information, the acquisition of voice data is simpler and more convenient. As a result, many techniques based on the analysis of speech and vocal patterns have been developed to diagnose Parkinson's disease. Designing machine learning models for analyzing voice signals to diagnose PD is considered critical to identify Parkinson's disease as early as possible. For example, studying voice patterns have become a great topic in PD diagnosis. Naranjo et all [2] designed a PD-diagnosed model from sound signals. Haq, Amin Ul, et al [3] suggested a PD prediction system based on machine learning. An L1-norm SVM-based feature was applied to classify the input signals into health and PD. Lahmiri et al. [4] investigated changes in voice patterns to diagnose PD. Statistical metrics including Mann-Whitney-Wilcoxon, a genetic algorithm was employed to select the most powerful features from voice signals. An SVM was used to classify the selected features with an accuracy of t (92.21%). Mustafa et al. [5] suggested a Multiple Feature Evaluation Approach based on machine learning algorithms to diagnose PD. Sabeena et al. [6] compared various machine learning methods for the diagnosis and classification of Parkinson's disease. Tai Yu Chen et al [7] proposed a supervised learning-based model to identify the patterns connected to PD patients' voices. A total of 1400 voice recordings were used to evaluate the proposed model. Jeancolas et all [8] used machine learning models to classify voice data. They acquired data from 256 French speakers including 117 participants with early PD, 41 with RBD, and 98 healthy. They achieved an accuracy of 89%.

Deep learning-based approaches have been used in feature classification and feature extraction for DP diagnosis. Ali, Liaqat, et al [9] combined linear discriminant analysis (LDA), genetic algorithm (GA), and neural networks (NN) in one model. They obtained an accuracy of 95% for differentiating between PD patients and healthy people. Gunduz, Hakan, et al [10] introduced a convolutional neural network model based on a set of vocals. Xu, Zhi-Jing, et al [11] used spectrogram deep convolutional generative adversarial network voice signal analysis. Onur Karaman et al.[12] employed a convolution neural network to classify the time-frequency characteristics of sound data, with an accuracy of 89.75%. shalin et al [13] Long-term and short-term memory networks have been assumed to predict Parkinson's disease. Based on the related works, although different methods of research have achieved good results on different datasets, it still lacks a robust method that is suitable for different datasets. This paper proposes a method that could be applied to most Parkinson's sound datasets.

2. Data Description

The PD database contains training and test files. The subjects of the training data are 20 Persons with Parkinson's (PWP) (6 females and 14 males) and 20 healthy persons (10 females and 10 males) who appealed at the Department of Neurology at Istanbul University's Cerrahpasa Faculty of Medicine. Numerous sound recordings (26 voice samples comprising words, sustained vowels, small sentences, and numbers) were taken from all participants. This dataset also includes each patient's UPDRS (Unified Parkinson's Disease Rating Scale) score, determined by knowledgeable medical professionals; patients in the test group have had Parkinson's disease for 0 to 6 years. Individual ages range from 43 to 77 and 45 to 83 for the test and control groups, respectively. A Trust MC-1500 microphone with a frequency response of 50 Hz to 13 kHz is used for recording. The patient is instructed to read or repeat the chosen text. At the same time, the microphone is tuned to 96 kHz, 30 dB, and is situated 10 cm away from the subject. After gathering the initial dataset of various sound recordings and running our experiments following the results, we acquired a separate test set from PWP using the same physician's examination procedure under the same situations to create the 168 recordings of this dataset, 28 people with Parkinson's disease were requested to speak the sustained vowels"o" and “three times each. Patients in the test group range in age from 39 to 79 and have had Parkinson's disease for 0 to 13 years. To verify the outcomes of our multiple sound recordings dataset, as an independent test set, we used

this dataset [14]; the UCI archives' patient database, accessible at **https://archive.ics.uci.edu/ml/datasets/Parkinson+Speech+Dataset+with++Multiple+Types+of+Sound+Recordings#**, was used in the experiments [15]

3. Materials and methods

3.1. Proposed Parkinson's disease detection framework

In this paper, an intelligent model for PD detection is suggested based on empirical mode decomposition (EMD). Figure 1 depicts the suggested Parkinson's disease detection framework. Where first, noise and other artifacts are eliminated through preprocessing of the raw data. Then, the filtered signals are partitioned into segments. A speech signal is made up of non-oscillatory transients and persistent oscillations. We extract entropy properties from sound recordings and apply them to several subsequent classification algorithms in this study. First, the input signals are passed through EMD to obtain a collection of IMFs, and then a set of entropy features is extracted from each IMFs. Finally, we aggregated all EMD features and used them to train the KNN classifier to detect PD. The extracted features are also sent to other classifiers for comparisons purpose

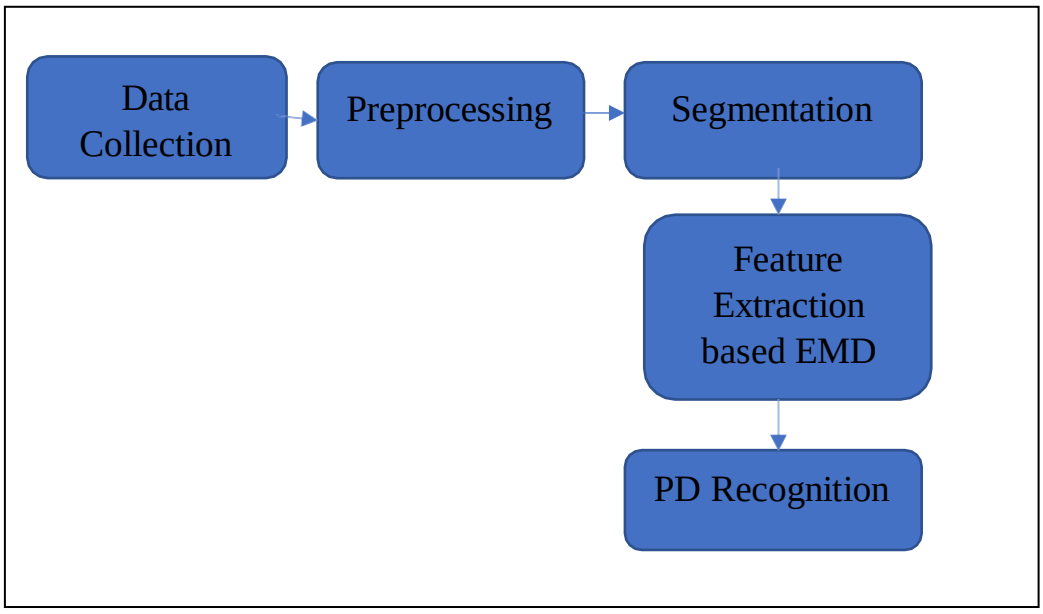

Figure 1. Proposed Parkinson's disease detection framework.

3.2 Pre-processing and segmentation

Voice signals were passed through a high pass filter and low pass filter to remove noises and improve voice signals quality. In this study, a high pass filter of 0.5 Hz cut off

frequency, and low pass filter of 70 Hz cut off frequency is applied to filter voice signals. After filtering the signals, voice signals were segmented into windows of 250 second length. The length of the window is set based on previous studies [16]

3.3. Empirical Mode Decomposition

The empirical mode decomposition (EMD) technique was created for the time-frequency analysis of real-world signals. It decomposes the input signal into a set of oscillatory modes known as intrinsic mode functions (IMFs). Then it applies the Hilbert transform to these intrinsic mode functions to generate meaningful instantaneous frequency estimations. EMD is suited for studying nonlinear and nonstationary signals [17,18,19], since the IMFs are produced directly from the data with no a priori assumptions about the data nature. EMD-based time-frequency analysis provides an extensive range of applications in signal processing and related domains [19]. As mentioned Real-valued EMD attempts to decompose a signal adaptively into a limited collection of oscillatory components known as "intrinsic mode functions" (IMFs) [20]. This is achieved by extracting quickly oscillating components (dominant modes) from the data and then deleting fewer dominant modes repeatedly (slowly oscillating components). The local mean of a signal is defined as these slowly oscillating components. After all of the dominant modes are removed, an IMF has a local mean value of zero, and the sifting process is stopped. More formally, if represents a real-valued input signal, then EMD produces a collection of IMFs denoted by [18,19, 21].

$$S(t) = \sum_{m=1}^{M} imf_{m(t)+r_m(t)} \qquad (1)$$

where $r_m\ (t)$ refers to the final residue, M denotes the number of IMFs.

The voice signal is one of the numerous fields in which it has been widely applied. In this study, we applied EMD to separate the voice signal into different frequency components to reveal the most important information [17,18]. As mentioned, the EMD approach decomposes non-stationary and nonlinear signal X into a set of IMFs. The decomposition does not require any kernel function, and the IMFs are just dependent on the signal itself. Because of this feature, the EMD is widely employed in a variety of fields [18,21,22].

IMFs are functions that have the following two properties:

1. The absolute difference between the total number of zero crossing points and a total number of function extrema is less than or equal to one.
2. The area under the curve created by the function's average lower and upper envelopes within any two consecutive extrema (one local maximum and one local minimum) is zero.

As a result, each IMF has lower frequency oscillations than the one just extracted. The EMD is a completely data-driven approach that does not require a pre-determined filter. The signal X(t) must have at least two extrema, one minimum, and one maximum, to be correctly decomposed into IMFs [20].

The process of extracting IMFS with EMD is described as figure.2.

Input: Voice signals S(t)

Output: A series of IMFs ($imfs_m$(t)) and a residual (r_m, (t))

Step1: Find all the local upper and the local lower values of voice signal S(t);

Step2: Construct the upper envelop E_u (t) and the lower envelop E_l() by using a method known as

cubic spline interpolation;

Step3: Calculate the mean of envelop: M(t) (E_u (t)+ E_l(t))/2;

Step4: Subtract the mean M(t) from the original voice signal S(t), then get a

new signal d(t)=S(t)-M(t);

Step5: if the signal d(t) satisfies the two properties of IMFs

$imf_{m(t)}$ =d(t);

Else

d(t) as the original signal and repeat steps 1 to 5

Step6: S(t)=S(t)- imf_m(t)

If S(t) is a monotonic function

r_m (t) =S(t) and the algorithm stop

Else

Figure 2. The process of extracting IMFS with EMD [16]

3.4. Entropy Features

Entropy is a term used to quantify the amount of chaos or randomness in a signal [23]. The greater the entropy, the more irregular the times series. To analyze the spectrum complexity in the supplied time series data, spectral entropy metrics such as Shannon entropy and log-energy entropy are used among different classes of entropy [18]. Also, Entropy is defined in thermodynamics as a measure of the degree of the disorder [24]. Entropy is an indication of the amount of information contained in a broader probability distribution in the setting of information theory, also Entropy has been defined as a measure of the time series' complexity [25].

In this work, we used three types of entropy features

Dispersion Entropy, Approximate entropy, and Shannon entropy

3.4.1 Shannon entropy (ShEn):

Entropy is calculated by multiplying the power level at each frequency by the logarithm of the power level's inverse. Finally, the spectral entropy of the time series is calculated using the method below [24].

$$ShEn = \sum_{f} p_f \log\left(\frac{1}{p_f}\right) \qquad (2)$$

the Shannon entropy, which calculates the mean value of the signal's information, suppose a voice signal $x = [x_1, x_2, x_3, \dots, x_n]$with a probability distribution function $p(x_i)$, the Shannon entropy is computed as:

$$\text{Shanon}_{\text{entropy}} = -\sum_{i=1}^{m} p_i^2 \log_2 (p_i^2) \qquad (3)$$

$$Shanon_{entropy} = \sum_{i=1}^{m} p_i^2 log_2 \qquad (4)$$

3.4.2 Approximate entropy (ApEn):

That measures a regularity of a time series [18]. ApEn (approximate entropy) is a time series complexity metric. It is widely employed in a wide range of biological signal processing applications, including EEG epileptic activity analysis, background activity, coronary artery disease (CAD) heart rate signal analysis, and so on. It assesses a time series' randomness or regularity in various dimensions [25].

approximate entropy (ApEn) that measures a regularity of a time series. the ApEN is calculated using the following formula.

$$\mathrm{ApEn}\,(z, r, L) = \boldsymbol{\phi}^{z}(r) - \boldsymbol{\phi}^{z+1}(r) \qquad (5)$$

Where $\phi^{z}(r)$ is calculated using the following formula:

$$\phi^{z}(r) = \frac{1}{L - z + 1} \sum_{i} \ln(C_i^z(r)) \qquad (6)$$

and C_i^z is the correlation integral given by

$$C_i^z(r) = \frac{1}{L - z + 1} L_i^r \quad i = 1, 2, \ldots, L - z + 1 \qquad (7)$$

3.4.3 Dispersion Entropy (DisEn):

DisEn is a technique for characterizing time series irregularity derived from Shannon Entropy and symbolic dynamics for constructing a fast and robust algorithm to quantify the degree of irregularity of a signal segment under investigation. The approach entails representing the dynamics of a signal segment using a dispersion pattern distribution. Dispersion patterns are symbol sequences formed by the relative amplitude of samples from a whole signal segment quantized with a particular mapping function. As a result, when a signal segment can be represented with a small proportion of dispersion patterns, it has a low DisEn value compared to one that needs all conceivable dispersion patterns to be present in equal probability [26].

The following parameters are set during the implementation of a DisEn algorithm:

- **Embedding dimension:** the number of samples of each dispersion pattern utilized to represent the signal segment.
- **The number of classes:** the number of possible values for each sample in the dispersion pattern.
- **Mapping approach (logarithm sigmoid function):** which distributes the values of the dispersion pattern classes along the amplitude range of the investigated signal.
- **Time delay:** This can be used to add a lag in the algorithm's analysis of samples.

3.5. Classifiers

3.5.1 Least Squares-Support Vector Machine (LS-SVM)

Swinkez and Vandal developed the Least Squares Support Vector Machine (LS-SVM) in 1999 [27]. Instead of the traditional SVM method's second-order programming technique, A set of linear equations is used in this model.

The LS-SVM approach may solve classification and regression issues [28,29], where the support vector machine (SVM) is based on maximizing the distance between two hyperplanes. It is capable of generalization. The least squares support vector machine (LS-SVM) technique looks for the best hyperplane in the higher dimension input space to build a decision boundary between two separate sets of patterns. This technique was originally designed as a linear classifier. The SVM method has been expanded as the LS-SVM algorithm, the least squares variant of SVM [30].

In this paragraph, we show a least squares version of the SVM classifier by formulating the classification problem as:

$$\min_{w,b,c} s(w,b,c) = \frac{1}{2} w^T w + r\frac{1}{2}\sum_{k=1}^{N} e_k^2 \quad (8)$$

subject to the equality constraints

$$y_k[w^T \phi(x_k) + b] = 1 - e_k \ , k = 1, \ldots, N. \quad (9)$$

The Lagrangian is defined by one

$$L_{3(w,b,c;a)} = L_{3\,(w,b,c)} - \sum_{k=1}^{N} a_k\{ y_k[w^T\phi(x_k) + b] - 1 + e_k \} \quad (10)$$

where a_k are Lagrange multipliers (which, because of the equality restrictions, can now be either positive or negative).

The conditions for optimality

$$\frac{\partial L_3}{\partial w} = 0 \to w = \sum_{k=1}^{N} \alpha_k y_k \phi(x_k)$$
$$\frac{\partial L_3}{\partial b_3} = 0 \to \sum_{k=1}^{N} \alpha_k y_k = 0 \qquad (11)$$
$$\frac{\partial L_3}{\partial c_k} = 0 \to \alpha_k = r_{ek}, k = 1, \dots, N$$
$$\frac{\partial L_3}{\partial \alpha_k} = 0 \to y_k [w^T \phi(x_k) + b] - 1 + e_k = 0, k = 1, \dots, N$$

may be represented as the solution to the following set of linear equations directly.

$$\begin{bmatrix} I & 0 & 0 & -Z^T & w & 0 \\ 0 & 0 & 0 & -Y^T & b & 0 \\ 0 & 0 & \gamma I & -I & c & 0 \\ Z & Y & I & 0 & \alpha & 1 \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}, \qquad (12)$$

Where

$Z = [\phi(x_1)^T y_1; \dots; \phi(x_N)^T y_N], Y = [y_1; \dots; y_N], \vec{1} = [1; \dots; 1], c = [c_1; \dots; c_N], \alpha = [\alpha_1; \dots; \alpha_N]$ The solution is also provided by

$$\begin{bmatrix} 0 & -Y^T \\ Y & ZZ^T + r^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \qquad (13)$$

Mercer's condition may be applied to the matrix $\Omega = ZZ^T$once again.

$$\Omega_{kl} = y_k y_l \phi(x_N)^T \phi(x_{kl}) \qquad (14)$$

$$= y_k y_l \Psi(x_k, x_l) \qquad (15)$$

3.5.2 Support Vector Machine (SVM)

Vapnik introduced SVM in 1980′ [31], A supervised approach known as a Support Vector Machine (SVM) [4] classifies data and also succeeds successfully in regression. To split data into class labels, the SVM builds hyperplanes. It is mainly employed for binary classification issues, such as our study, where class 1 represents a Parkinson's disease patient and class 0 represents a healthy individual [32,33]. In binary classification, A high separation between two classes is achieved via the hyper-plane with the largest distance to the nearest training data point of a certain class. [33].

$$\gamma_i = \{1,0\} \quad (16)$$

where γ_i is the collection of binary class labels.

Therefore

$$x_i.w + c \geq +1 \qquad \gamma_i = 1 \quad (17)$$

$$x_i.w + c \geq -1 \qquad \gamma_i = 0 \quad (18)$$

where, x_i is the ith sample, c represents the bias, where w represents the normal of the hyperplane, and This is the hyperplane's distance from the origin.

The SVM technique is based on the notion of maximizing the hyperplane margin, which classifies the data optimally. Figure (3) shows an SVM classifier trained to segregate data points across N-dimensional space employing support vectors (SV). which are important data points located on the margin borders. In this case, N = 2 characteristics (x1, x2).

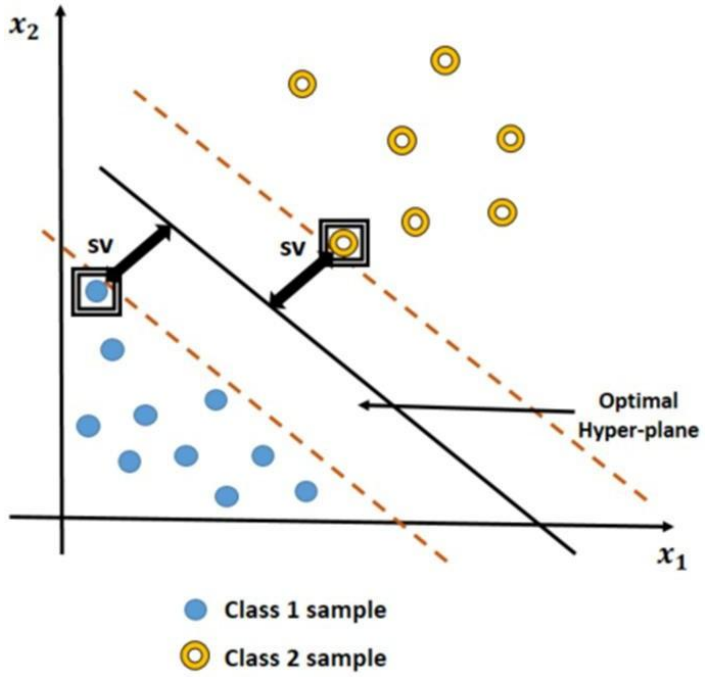

Figure 3. An SVM classifier with the best hyperplane [28].

If the data points can be separated linearly, then the data may be separated by an infinite number of hyperplanes. SVM seeks the linear function with the greatest margin discriminating between class labels. In the SVM, Input x is mapped onto a higher dimensional feature space using kernel functions. In this space, a linear model is then constructed. SVM kernel functions include linear, radial basis function (RBF), sigmoid and polynomial. The kernel functions project the n samples onto a new m-dimensional space. The SVM characteristics are determined by the kernel function utilized [32].

3.5.3 Bagged trees

Ensemble approaches (methods that aggregate the predictions of many models) first appeared in the 1990s. Leo Breiman invented bagging, short for bootstrap aggregation, as one of the earliest ensemble techniques. [34,35]. Bagging is a broad way to construct an ensemble that combines bootstrapping with any regression (or classification) model. The structure of the approach is quite straightforward, consisting of the stages in Algorithm (1). Following that, the ensemble models are used to forecast the new sample, and the k estimates are averaged to predict the bagging model.

Algorithm (1): Bagging ensemble pseudocode

Input:

dataset S= $\{(x_1, y_1), (x_2, y_2), , \ldots . . (x_n, y_n)\}$

Base learning algorithm L;

Number of base learning k.

Process

1. For I=1…, k;
2. $S_k = Bootstrap(S)$; % $Generate\ a\ bootstrap\ sample\ from\ S$
3. $h_{k=}L(S)$ % $Train\ a\ base\ learner\ h_k from\ the\ bootstap\ sampl$
4. end
5. **Output**:

H(x)=$arg_{y\in Y} max \sum_{k=1}^{k} l\,(h_k(S) = y)$ % $the\ value\ of\ l(a) is 1\ if\ a\ \ is\ true$ $0\ otherwise$.

3.5.4 K-Nearest-Neighbour (KNN) Classifiers

K-NN is a supervised classification technique, and KNN is a non-parametric approach for solving both regression and classification problems. This technique is one of the simplest and oldest for pattern recognition. It is used in machine learning, text categorization, data mining, Etc. It frequently produces efficient performance and, in certain situations, exceeds state-of-the-art classifiers in accuracy [36]. using the K-closest samples from the training set, the technique predicts a new sample (similar to Fig. 4) [31,36]. Each observation in a learning set is represented as a point in an n-dimensional space, where n is the number of predictor variables. We seek the K points that are closest to this pattern to predict the class of an observation. The target variable's class has the most representation among the K's closest neighbors. The performance of a KNN classifier is also determined by the "d" distances between the tested and training data. Manhattan, Euclidian, and Chebyshev distances are the most commonly used examples [37]

Furthermore, the most generally used metric is Euclidean distance (i.e., the straight-line distance between two samples) [25], which is defined as follows:

$$\sum_{j=1}^{p}\left(\left(x_{1j}-x_{2j}\right)^{2}\right)^{\frac{1}{2}} \quad (19)$$

where $x_1$and x_2 are two different examples. Minkowski distance, an extension of Euclidean distance, is defined as follows:

$$\sum_{j=1}^{p}\left(\left(x_{1j}-x_{2j}\right)^{q}\right)^{\frac{1}{q}} \quad (20)$$

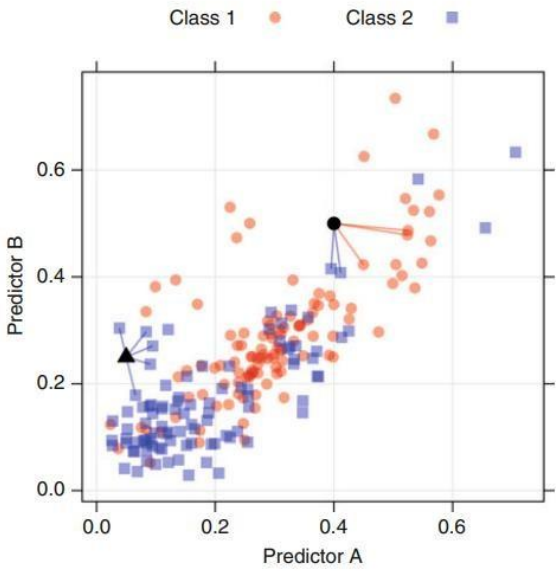

Figure 4. The categorization model of K-nearest neighbors [35].

Because it completely depends on every sample in the training set, it has limitations such as time complexity and memory requirements [36].

Algorithm 2: The pseudocode for KNN is as follows

1. Load the data
2. select K value
3. For each of the data's data points:
   - To each training data sample, calculate the Euclidean distance.
   - Sort and keep the ordered list of distances.
     - select the top K entries from the sorted list
     - Label the test point using the classes that compose the majority of the chosen points.

Figure 5. KNN pseudocode [26]

3.5.5 K means clustering

J.B. MacQueen proposed the K-Means method, which is based on dividing and is a type of cluster algorithm. This unsupervised approach is commonly used in pattern recognition and data mining.

The primary concept behind these methods is this algorithm's basis is the square error and error criterion, which seeks to reduce the cluster performance index. To attain the best possible outcome, this strategy aims to locate K divisions that meet a specified condition. Firstly, choose several dots to represent the first cluster focus points. (Typically, the first K sample dots with revenue are chosen to represent the first cluster focus point.); Secondly, we will use the minimum distance criterion to gather the remainder of the dots to their focal points, and then we will get the initial classification; if the classification is unreasonable, we will modify it (calculate each cluster focal point again), and we will iterate until we get a satisfactory classification. The division-based K-Means technique is a sort of cluster algorithm that has the advantages of being concise, efficient, and quick. [37].

Furthermore, this method is heavily dependent on starting dots and sample selection differences, which always results in various conclusions. Furthermore, the gradient method is always used to achieve the extremum in this goal function-based approach. Because the gradient approach constantly searches in the direction of energy drop, if the initial cluster focus point is erroneous, the entire process will sink toward the local minimum point.

3.6 Performance metrics

The five-performance metrics accuracy (Acc), specificity (Spec), sensitivity (Sen), positive predictive value (PPV), and Fscore are used to evaluate the performance of the classifiers. The formulas to calculate the aforementioned performance metrics are provided in Table 1:

Table 1. Performance metrics for the classifier.

parameter	Expression
Accuracy	$\text{Acc (\%)} == \frac{TP + TN}{TP + FP + TN + FN} * 100\%$
Sensitivity	$\text{Sen (\%)} = \frac{TP}{TP + FN} * 100\%$
Specificity	$\text{Spec (\%)} = \frac{TN}{TN + FP} * 100\%$
Precision	$\text{Pre (\%)} = \frac{TP}{TP + FP} * 100\%$
F-score	$\text{F-score (\%)} = 2 * \frac{* \text{Precision} * \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} * 100\%$

4. Experimental results and discussion

In this paper, an intelligent mode for PD disease detection is designed. The proposed model is implemented in MATLAB 2020A. The proposed model was evaluated on a publicly available UCI dataset. The proposed model is based on EMD to classify the dataset into healthy and unhealthy subjects. Firstly, voice signals were preprocessed to reduce noise and outliers. The signals were then passed through the EMD to decompose the signals into veracious IMFs. Then, several features were extracted to represent the voice signals.

With EMD, where Figure (6) shows the exemplary patient and healthy voice data and their respective three-level IMFs using EMD. Through EMD, the high-frequency content is first obtained from the voice data, and the frequency of remaining IMFs gradually decreases. Hence, to identify the localized features in the IMF, from the IMFs, the entropy features such as Shannon entropy, Dispersion entropy, and Approximate entropy are obtained.The dataset was divided randomly into 20% for the testing and 80% for the training.

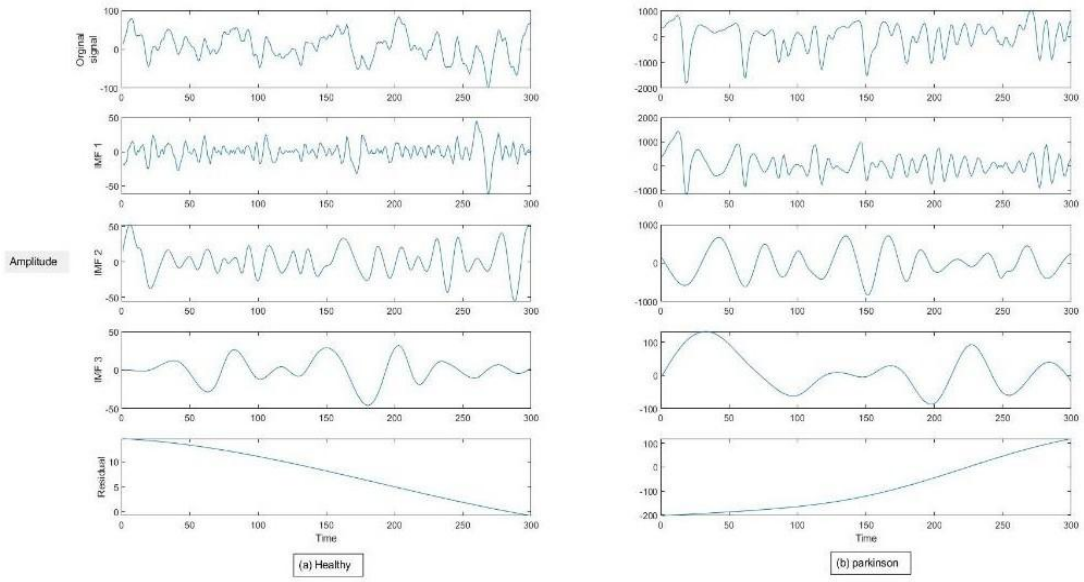

Figure 6: EMD of (a) Healthy (b) patients.

4.1 Classifiers Performance Based on Empirical Mode Decomposition (EMD).

In this experiment, three Entropy features from each IMFS of EMD were used to form the PD model. The voice recording signals were randomly partitioned into 80% for training and 20% for testing. The extracted features were sent to the K-nearest, LS-SVM, SVM, Bagged tree, and Kmeans.

Table 2 reports the classification results based on Empirical mode decomposition (EMD)

Table 2 Classification Results Based on Empirical Mode Decomposition (EMD)

Classifier	Accuracy	Sensitivity	Specificity	Precision	F1score
KNN	0.97	0.95	1	1	0.97
bagged tree	0.97	0.94	1	1	0.97
LS_SVM	0.92	0.90	0.94	0.95	0.93
SVM	0.92	0.90	0.94	0.95	0.93
Kmeans	0.84	1	0.68	0.75	0.86

4.2 Cross-Fold Validation

Ten folds or groups were created from the PD Speech data set. The remaining k-1 group served as training data, whereas one group served as test data. Following that, the classifier was evaluated and scores were stored. The procedure was carried out k times, using various folds of test data each time. Combining all of the scores produced the classier's final score, which is represented by the equation below:

$$M = \frac{1}{r} \times \sum_{n=1}^{r} A_n \pm \sqrt{\frac{\sum_{n=1}^{r} (A_n - \bar{A})^2}{r-1}} \qquad (21)$$

Here, M designates the overall performance metric for that classifier, while A_n stands for the nth ($1 \leq n \leq K$) fold performance metric [37].

The results of each fold are then averaged to provide an overall estimate of the model's performance. The results of the cross-validation were presented in Table 3, which shows the result for each classifier, as well as the average values across all folds. This approach provides a robust estimate of the model's performance and helps to reduce the risk of overfitting the training data.

TABLE 3.TEN CROSS-VALIDATION

Classifier	Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
LS_SM	0.95	0.95	1	1	1	1	1	1	1	1
SVM	1	1	1	1	1	1	0.90	1	1	1
KNN	1	1	1	0.95	1	1	1	1	0.90	1
Bagged	1	1	1	1	1	1	1	0.90	1	1

4.3 Complexity time of the proposed model

We examine the temporal complexity of the algorithms used to optimize the Parkinson's disease detection system in this section. The time complexity of the algorithms used in the detection system for Parkinson's disease was thoroughly analyzed, and the results indicate that the system is efficient and effective. The percentages obtained from the data showed that the time complexity was low, indicating that the system is less time-consuming, which is a desirable feature. With such a low time complexity, the proposed approach can significantly improve the classification accuracy of the system. the results of the time complexity are shown in Figure (7).

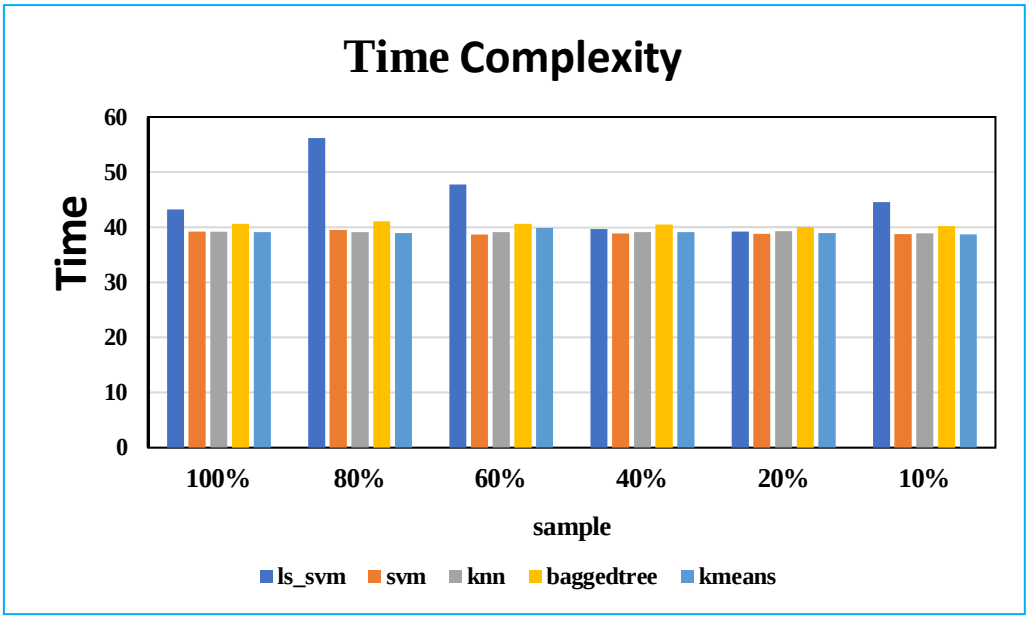

Figure 7. The proposed method's time complexity. Time (seconds)

6-Conclusion

Based on EMD feature extraction methods, this paper has presented a framework for the detection of Parkinson's disease. The PD dataset from the UCI database is used for testing. we compute the entropy, Shannon and approximation-based features for ML classification after extracting the IMFs using the EMD approach. A ten-fold cross validation method is used to reduce the overfitting problem in ML algorithms. And the confusion matrix has been used to assess the performance of the four classifier models, SVM, KNN, bagged tree and k _means. The experimental results demonstrate the potential of the suggested framework employing the KNN technique a suitable classification accuracy of 97% was achieved. The suggested method can be used in future research to classify other complex diseases across a larger dataset.

REFERENCES

[1] **Poorjam, Amir Hossein, et al**. "Automatic quality control and enhancement for voicebased remote Parkinson’s disease detection." Speech Communication 127 (2021):

[2] **Naranjo, Lizbeth, et al.** "A two-stage variable selection and classification approach for Parkinson’s disease detection by using voice recording

replications." Computer methods and programs in biomedicine 142 (2017): 147-156.

[3] Haq, Amin Ul, et al. "Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings." IEEE access 7 (2019): 37718-37734.

[4] **Lahmiri, Salim, and Amir Shmuel.** "Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine." Biomedical Signal Processing and Control 49 (2019): 427-433.

[5] Mostafa, Salama A., et al. "Examining multiple feature evaluation and classification methods for improving the diagnosis of Parkinson's disease." Cognitive Systems Research 54 (2019): 90-99.

[6] **Sabeena, B., S. Sivakumari, and P. Amudha**. "A technical survey on various machine learning approaches for Parkinson's disease classification." Materials Today: Proceedings (2020).

[7] **Tai, Yu Chen, et al.** "A voice analysis approach for recognizing Parkinson's disease patterns." IFAC-PapersOnLine 54.15 (2021): 382-387.

[8] **Laetitia, et al.** "Voice characteristics from isolated rapid eye movement sleep behavior disorder to early Parkinson's disease." Parkinsonism & Related Disorders 95 (2022): 86-91

[9] **Ali, Liaqat, et al.** "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network." IEEE journal of translational engineering in health and medicine 7 (2019): 1-10.

[10] **Gunduz, Hakan.** "Deep learning-based Parkinson's disease classification using vocal feature sets." IEEE Access 7 (2019): 115540-115551.

[11] **Xu, Zhi-Jing, et al.** "Parkinson's disease detection based on spectrogram-deep convolutional generative adversarial network sample augmentation." IEEE Access 8 (2020): 206888-20690.

[12] **Karaman, Onur, et al.** "Robust automated Parkinson disease detection based on voice signals with transfer learning." Expert Systems with Applications 178 (2021): 115013.

[13] **Shalin, Gaurav, et al.** "Prediction and detection of freezing of gait in

Parkinson's disease from plantar pressure data using long short-term memory neural-networks." *Journal of neuroengineering and rehabilitation* 18.1 (2021): 1-15.

[14] Sakar, Betul Erdogdu, et al. "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings." *IEEE Journal of Biomedical and Health Informatics* 17.4 (2013): 828-834.

[15] **Wrobel, Krzysztof.** "Diagnosing Parkinson's disease by means of ensemble classification of patients' voice samples." Procedia Computer Science 192 (2021): 3905-3914.

[16] **Diykh, M., Miften, F.S., Abdulla, S., Deo, R.C., Siuly, S., Green, J.H. and Oudahb, A.Y**., 2022. Texture analysis-based graph approach for automatic detection of neonatal seizure from multi-channel EEG signals. Measurement. 190, p.110731.

[17] **Zhang, Tao, et al.** "Parkinson disease detection using energy direction features based on EMD from voice signal." Biocybernetics and Biomedical Engineering 41.1 (2021): 127-141.

[18] **Leung, Alexander Kai-man, Foo-tim Chau, and Jun-bin Gao.** "A review on applications of wavelet transform techniques in chemical analysis: 1989–1997." Chemometrics and Intelligent Laboratory Systems 43.1-2 (1998): 165-184.

[19] **Tzanetakis, George, Georg Essl, and Perry Cook**. "Audio analysis using the discrete wavelet transform." Proc. conf. in acoustics and music theory applications. Vol. 66. 2001.

[20] **ur, Rehman, Naveed, and Danilo P. Mandic**. "Empirical mode decomposition for trivariate signals." IEEE Transactions on signal processing 58.3 (2009): 1059-1068.

[21] **Roberts, Andrew**. "Guide to WEKA." (2005).

[22] **Huang, Zuo, and Bingo Wing-Kuen Ling.** "Joint ensemble empirical mode decomposition and tunable Q factor wavelet transform based sleep stage classifications." Biomedical Signal Processing and Control 77 (2022): 103760.

[23] **Jain, A. K., Duin, R. P. W., & Mao, J. (2000).** Statistical pattern recognition: A review. IEEE Transactions on pattern analysis and machine

intelligence, 22(1), 4-37.

[24] **Finney, Ross L., and Maurice D. Weir**. Thomas' calculus: early transcendentals. Addison-Wesley, 2001.

[25] **Kannathal, N., et al.** "Entropies for detection of epilepsy in EEG." Computer methods and programs in biomedicine 80.3 (2005): 187-194.

[26] **Li, Yan, and Peng Paul Wen**. "Clustering technique-based least square support vector machine for EEG signal classification." Computer methods and programs in biomedicine 104.3 (2011): 358-372.

[27] **Hoq, Muntasir, Mohammed Nazim Uddin, and Seung-Bo Park.** "Vocal feature extraction-based artificial intelligent model for Parkinson's disease detection." Diagnostics 11.6 (2021): 1076.

[28] **Goberman, Alexander M**. "Correlation between acoustic speech characteristics and non-speech motor performance in Parkinson disease." Medical science monitor 11.3 (2005): CR109-CR116.

[29] **Kannathal, N., et al.** "Entropies for detection of epilepsy in EEG." Computer methods and programs in biomedicine 80.3 (2005): 187-194.

[30] **Li, Yan, and Peng Paul Wen.** "Clustering technique-based least square support vector machine for EEG signal classification." Computer methods and programs in biomedicine 104.3 (2011): 358-372.

[31] **Kunhimangalam, Reeda, Sujith Ovallath, and Paul K. Joseph**. "A novel fuzzy expert system for the identification of severity of carpal tunnel syndrome." BioMed research international 2013 (2013).

[32] **Meghraoui, Djamila, et al.** "A novel pre-processing technique in pathologic voice detection: Application to Parkinson's disease phonation." Biomedical Signal Processing and Control 68 (2021): 102604.

[33] **Eskidere, Ömer, Figen Ertaş, and Cemal Hanilçi**. "A comparison of regression methods for remote tracking of Parkinson's disease progression." Expert Systems with Applications 39.5 (2012): 5523-5528.

[34] **Kuhn, Max, and Kjell Johnson**. Applied predictive modeling. Vol. 26. New York: Springer, 2013.

[35] **Breiman, Leo.** "Bagging predictors." Machine learning 24 (1996): 123-140.

[36] **Hassanat, Ahmad Basheer, et al**. "Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach." arXiv preprint arXiv:1409.0919 (2014).

[37] **Rybach, David, et al.** "Audio segmentation for speech recognition using segment features." 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2009.