Biomedical Engineering Letters (2023) 13:613–623
https://doi.org/10.1007/s13534-023-00283-x

ORIGINAL ARTICLE

Time-frequency analysis of speech signal using Chirplet transform for automatic diagnosis of Parkinson's disease

Pankaj Warule[1] · Siba Prasad Mishra[1] · Suman Deb[1]

Received: 5 January 2023 / Revised: 22 April 2023 / Accepted: 25 April 2023 / Published online: 8 May 2023

Abstract
Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder in the world after Alzheimer's disease. Early diagnosing PD is challenging as it evolved slowly, and its symptoms eventuate gradually. Recent studies have demonstrated that changes in speech may be utilized as an excellent biomarker for the early diagnosis of PD. In this study, we have proposed a Chirplet transform (CT) based novel approach for diagnosing PD using speech signals. We employed CT to get the time-frequency matrix (TFM) of each speech recording, and we extracted time-frequency based entropy (TFE) features from the TFM. The statistical analysis demonstrates that the TFE features reflect the changes in speech that occurs in the speech due to PD, hence can be used for classifying the PD and healthy control (HC) individuals. The effectiveness of the proposed framework is validated using the vowels and words from the PC-GITA database. The genetic algorithm is utilized to select the optimum features subset, while a support vector machine (SVM), decision tree (DT), K-Nearest Neighbor (KNN), and Naïve Bayes (NB) classifiers are employed for classification. The TFE features outperform the breathiness and Mel frequency cepstral coefficients (MFCC) features. The SVM classifier is most effective compared to other machine-learning classifiers. The highest classification accuracy rates of 98% and 99% are achieved using the vowel /a/ and word /atleta/, respectively. The results reveal that the proposed CT-based entropy features effectively diagnose PD using the speech of a person.

Keywords Chirplet transform · Genetic algorithm · Parkinson's disease (PD) · Time-frequency representation · Speech Pathology · Support vector machine

1 Introduction

Parkinson's disease (PD) is a neurological disease affecting around 1% of individuals over the age of 65 [1, 2]. PD exhibits both motor and non-motor symptoms. The motor symptoms of PD include tremors, bradykinesia, stiffness (rigidity), and postural instability. The most common non-motor symptoms are dysautonomia, discomfort, sensory dysfunction, and cognitive impairment [3]. The typical PD diagnosis process takes more than two years [4]. As a result, a new diagnostic method is needed to help with the PD diagnosis process based on the patient's symptoms. Approximately 89% of individuals suffering from PD exhibit speech impairments [5]. This speech impairment mainly includes fundamental frequency fluctuation, aperiodicity in vocal fold vibration, reduced speech intensity, hoarseness, and irregular articulation. Hence, speech is an excellent option for offering diagnostic clues for the automatic detection of PD.

Numerous studies have been conducted to detect PD using speech signals. The feature extraction and classification stages are the two key steps of the speech signal-based PD detection system. An appropriate speech processing algorithm is utilized to extract meaningful features from speech during the feature extraction. These features are applied to the machine learning classifier to evaluate the effectiveness of the system. Sakar et al. [6] developed a database for PD and healthy individuals' speech samples with sentences, words, and sustained vowels. They evaluated the database by extracting 26 acoustic features and machine learning classifiers. Tsanas et al. [7] used 132 dysphonia measures extracted from sustained vowels.

✉ Pankaj Warule
d20ec007@eced.svnit.ac.in

Siba Prasad Mishra
ds20ec005@eced.svnit.ac.in

Suman Deb
sumandeb@eced.svnit.ac.in

[1] Department of Electronics Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, India

Then, using feature selection methods, an efficient feature set was chosen and applied to machine learning classifiers to classify PD and healthy control (HC) individuals. Nilashi et al. [8] presented a machine-learning approach for PD progression. Based on 16 acoustic characteristics, they established a novel approach for predicting total unified Parkinson's disease rating scale (UPDRS) and motor UPDRS. Orozco-Arroyave et al. [9] examined the discriminating capability of various spectral and cepstral coefficients using voice recordings of 24 isolated words and the five Spanish vowels for classifying PD and HC individuals. Karan et al. [10] proposed instantaneous energy deviation cepstral coefficient (IEDCC) characteristics for classifying PD and healthy individuals using the Hilbert spectrum (HS). In another study, Karan et al. [11] analyzed the performance of time-frequency features derived from the decomposition of a time-frequency matrix (TFM) using non-negative matrix factorization for classifying PD and HC individuals. Narendra et al. [12] employed voice source information extracted using glottal features for classifying PD and HC individuals. They investigated the performance of traditional pipelines as well as the end-to-end classifier architectures. Karan et al. [13] proposed Hilbert cepstral coefficients (HCCs) features to analyze the voice tremor of PD patients utilizing a combined approach of VMD and HS. Mehmet et al. [14] employed variational mode decomposition (VMD) to denoise speech signal and spectrograms of the denoised signal applied to pre-trained convolutional neural networks (CNN) and long short-term memory (LSTM) to classify PD and HC individuals. Hirešet al. [15] presented an ensemble of CNNs trained with multiple fine-tuning to diagnose PD from voice recordings. Quan et al. [16] proposed a two-dimensional CNN to extract time series dynamic features and, subsequently, a one-dimensional CNN to capture the dependencies between these time series for the diagnosis of PD from the speech signal. Karan et al. [17] proposed an intrinsic mode function cepstral coefficient (IMFCC) based on empirical mode decomposition (EMD) to identify PD and healthy persons. Fujita et al. [18] introduced a recurrent neural network (RNN) with a hyperbolic secant gate for the diagnosis of PD using speech which achieved equivalent accuracy to LSTM and gated recurrent unit (GRU) with fewer hyperparameters. Goyal et al. [19] proposed a hybrid resonance-based sparse signal decomposition and time-frequency algorithm for feature extraction from speech recordings for diagnosing PD.

Researchers investigated numerous time-domain and frequency-domain features in the literature for classifying PD and HC individuals from speech signals, although time-frequency-based features are not employed extensively. The discontinuities and abrupt changes occur in the speech of PD patients as the patients are unable to regulate the limbs and muscles used to produce speech. Separate time and frequency analyses may not capable of capturing these changes in the speech of PD patients [11]. The time-frequency representation of signal models the temporal and frequency changes jointly; hence we have considered the hypothesis that the feature derived from the time-frequency representation of speech signal gives significant information for classifying PD and HC patients. This study employed the Chirplet transform (CT) to analyze speech recording and time-frequency domain feature extraction for classifying PD and HC individuals. The CT is effective for non-stationary signals [20, 21]. As speech is a non-stationary signal [22], it is expected that the time-frequency domain information obtained using CT of speech recording may provide useful information for classifying PD and HC individuals. In this work, we employed CT to get the TFM representation of each speech recording. Then, we used the TFM to get the time-frequency based entropy (TFE) features.

This paper follows the following structure: Sect. 2 discusses the proposed method for diagnosing PD using speech signals. Section 3 provides the experiment's findings and a discussion. The conclusion is presented in Sect. 4.

2 Proposed method

The proposed method for diagnosing PD using speech recording is illustrated in Fig. 1. It mainly consists of stages such as time-frequency analysis of speech recordings using CT, TFE feature extraction from the TFM, feature selection using a genetic algorithm, and SVM, DT, KNN, and NB classifiers for classifying PD and HC individuals.

2.1 Database

This study makes use of the PC-GITA database [1] to assess the proposed method. It includes Spanish speech samples of 50 HC and 50 PD individuals. Each PD and HC group has 25 men and women. The database has a balanced distribution of the speaker's gender and age. Men with PD range in

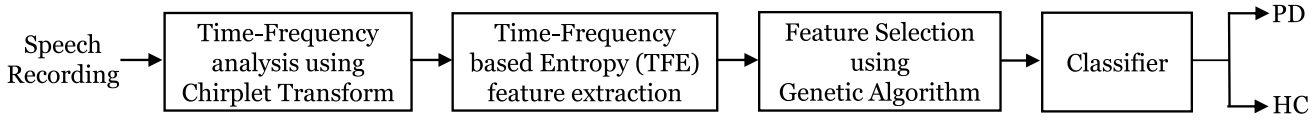

Fig. 1 The proposed method for diagnosing PD using speech recording

age from 33 to 77 years, and women from 44 to 75 years. For healthy individuals, the age ranges for men and women are 31–86 and 43–76 years, respectively. Three experienced phoneticians examined the participants' speech. In this study, we have evaluated the proposed method on the following speech task of the PC-GITA database: (1) three repetitions of each of the five sustained vowels (/a/, /e/, /i/, /o/, and /u/), (2) Reading the five words (/apto/, /atleta/, /braso/, /globo/, and /petaka/). The sampling rate of speech recordings is 44.1 kHz. Before processing the speech recordings, every recording is downsampled to 8 kHz.

2.2 Chirplet transform (CT)

The CT is utilized for the time-frequency domain analysis of non-stationary signals. For a speech signal $s(n)$ with length N, i.e., $n = 1, 2, 3, \ldots, N$, the CT is defined as [20, 21]:

$$C_{\alpha,\sigma}(k, n_0) = \sum_{n=1}^{N} z(n) e^{-\frac{2\pi k n}{N}} \Psi^*_{\tau,\alpha,\sigma}(n) \tag{1}$$

where $z(n)$ is the analytical signal of speech signal $s(n)$. It is given by

$$z(n) = s(n) + jH[s(n)] \tag{2}$$

where $H[s(n)]$ is the Hilbert transform of speech signal $s(n)$. The term $\Psi^*_{n_0,\alpha,\sigma}(n)$ is complex conjugate of window function $\Psi_{n_0,\alpha,\sigma}(n)$, given by

$$\Psi_{n_0,\alpha,\sigma}(n) = w_\sigma(n - n_0) e^{-j\frac{\alpha}{2}(n-n_0)^2} \tag{3}$$

where $w_\sigma(n - n_0)$ represents a Gaussian window function given by [20, 21]

$$w_\sigma(n) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{n^2}{2\sigma^2}} \tag{4}$$

The complex values of the TFM of speech signal $s(n)$ can be represented as

$$C_{\alpha,\sigma}(k, n_0) = C^R_{\alpha,\sigma}(k, n_0) + jC^I_{\alpha,\sigma}(k, n_0) \tag{5}$$

The magnitude of TFM of speech signal is given as

$$| C_{\alpha,\sigma}(k, n_0) | = \sqrt{\left[C^R_{\alpha,\sigma}(k, n_0)\right]^2 + \left[C^I_{\alpha,\sigma}(k, n_0)\right]^2} \tag{6}$$

In this study, we have examined the magnitude of the TFM for speech recording of various speech tasks (sustained vowels and isolated words) of PD and healthy individuals. For PD and healthy classes, the speech signals for sustained vowel /a/ are shown in Fig. 2a, b, and corresponding time-frequency contour plots are shown in Fig. 2c, d. The contour plots show significant visual alterations in the time-frequency properties of the speech signal as a result of PD. Hence, we may utilize the features derived from the TFM of speech for classifying PD and healthy individuals. We extracted the time-frequency based entropy (TFE) feature from the TFM in this work. The TFE feature is calculated

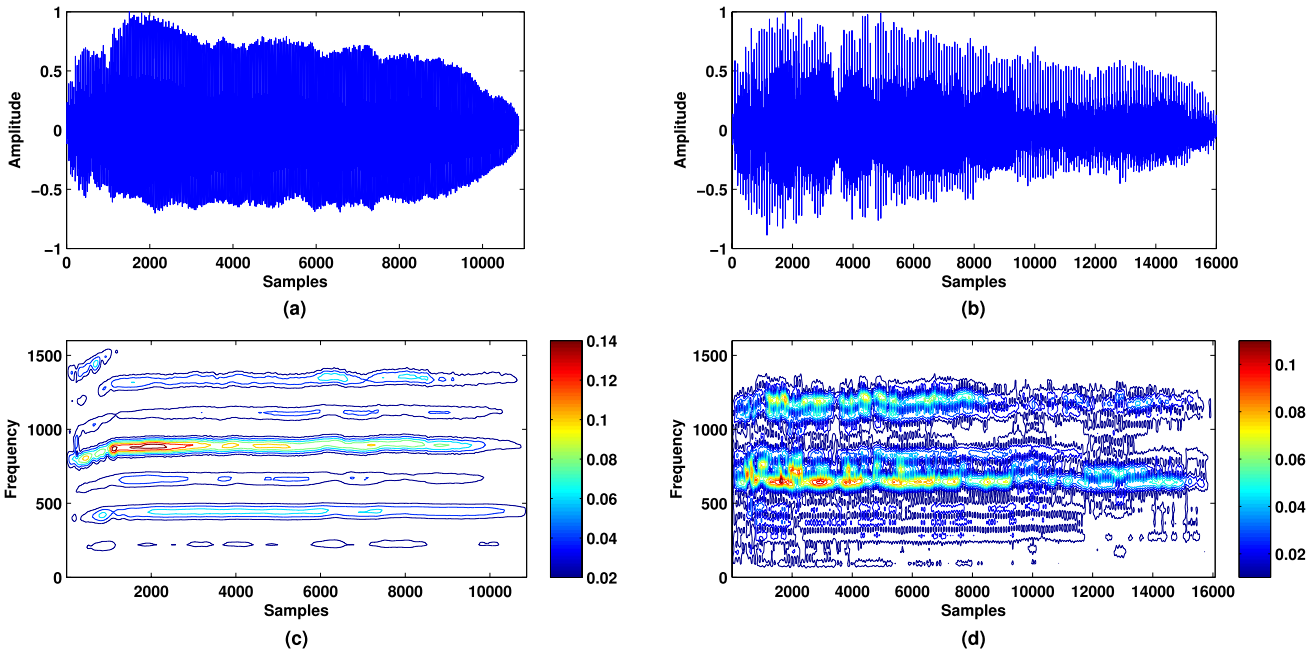

Fig. 2 Speech signals (vowel /a/) and their time-frequency contour plots **a** HC speech, **b** PD speech, **c** Time-frequency contour plot for HC speech, and **d** Time-frequency contour plot for PD speech

using the histogram of mth frequency component in the TFM representation. The mth frequency component's probability value is calculated as [23]

$$P_b(m) = \frac{h_b(m)}{\sum_{b=1}^{B} h_b(m)} \tag{7}$$

where $H_b(k)$ is the histogram of mth frequency component with B number of bins. The TFE features for mth frequency component is given by [24]

$$TFE_m = -\sum_{b=1}^{B} P_b(m) log_2 \left[P_b(m)\right] \tag{8}$$

In this work, we have extracted TFE features for all the frequency components to form 256 dimensional feature vector for each speech recording. In the next step, we have employed the genetic algorithm for feature selection. The reduced feature set obtained using genetic algorithm are employed for classifying PD and healthy individuals.

We have contrasted the efficacy of the proposed TFE features with the MFCC and breathiness features [25–29]. Which have often been used for speech pathology detection in the literature. Breathiness features are an important parameter for analyzing voice quality since it reflects information about speech perturbations [25]. These breathiness features include jitter, shimmer, number of voiced frames (NVF), harmonic energy (HE), harmonic-to-noise ratio (HNR), harmonic energy of residue (HER), and glottal-to-noise excitation ratio (GNER). In this work, we extracted these breathiness features for classifying PD and HC individuals as given in [25, 26]. We have also extracted 39-dimensional MFCC (13-MFCC, 13-ΔMFCC, 13-ΔΔ MFCC) features [30].

2.3 Feature selection

Feature selection is the most prevalent and widely employed data preprocessing technique for determining the optimal set of features. Finding the optimal subset of features involves a search in the space defined by all available feature combinations. This search might be laborious and time-consuming. This study employed a genetic algorithm to select optimal TFE, breathiness, and MFCC features for classifying PD and HC individuals.

2.3.1 Genetic algorithm

Evolutionary computation is a breakthrough innovation that helps to solve and optimize problems in the computing world [31]. A genetic algorithm is a flexible evolutionary computation based on genealogy, and natural phenomenology [32]. It is an important heuristic algorithm that imitates Darwin's theory of evolution [31]. Naturally, they are supposed to optimize the process by selecting the optimal answer and discarding the remainder [33].

The steps below summarize the execution of Genetic algorithm for feature optimization. The process begins by randomly creating a population of any size. The fitness function is used to determine how fit each unit of the population is to accomplish the solution by estimating their fitness value [34]. Then, we choose the features with the highest fitness values while discarding those with the lowest fitness values to preserve the fittest unit for the next generation and enhance the population's overall fitness. Two parents are chosen from each generation based on their fitness scores, and their vector entries are combined in the crossover phase for reproduction. Furthermore, offspring are formed from the parents by making random modifications to each parent alone, known as mutation. Using crossover, the algorithm finds the best genes from many parents and merges them to create potentially superior offspring. Even though the mutation rate is very low, it increases population diversity, increasing the likelihood that the algorithm would generate individuals with higher fitness values [34]. To create the next generation, the present population is replaced with offspring, then the procedure is repeated. The optimization will terminate when the number of generations reaches its pre-defined value.

2.4 Classifiers

The discrimination capability of proposed features is evaluated using SVM, DT, KNN, and NB classifiers along with a genetic algorithm.

2.4.1 Support vector machine (SVM)

The SVM is widely employed in the fields of speech pathology and emotion recognition [35–39]. For a binary classification task, it divides the two classes using hyperplanes as efficiently as possible. It determines decision boundaries by optimizing a certain mathematical function over a specific data set [40]. In addition, kernel functions are also employed to convert the initial feature space into a higher-dimensional space that it can be separated linearly [41]. The SVM uses convex optimization to find a solution that is globally optimum. In this study, we employed SVM classifier with radial basis function (RBF) for classifying speech recordings of PD and HC individuals.

2.4.2 Decision tree (DT)

A DT is a nonparametric supervised learning algorithm that can be employed for classification as well as regression problems. One distinguishing feature of DT among all

machine learning classifiers is that they are computationally inexpensive [42]. It comprises a tree-like, hierarchical structure with a root node, branches, internal nodes, and leaf nodes [43]. Each node in the tree represents an attribute test, each branch represents the test's outcome, and each leaf represents a class or class distribution [44]. A classifying rule starts with the root node at the top and ends with the leaf at the bottom.

2.4.3 K-nearest neighbor (K-NN)

The K-NN algorithm is a well-known supervised machine-learning approach that has been widely explored for classification, regression, data mining, and pattern recognition [45]. The k-nearest neighbor (k-NN) algorithm is a non-parametric instance-based approach that classifies a point by identifying the k vectors closest to it [46]. In this method, instead of creating a model, each new observation must be compared to the whole set of training data. Hence it is also known as lazy evaluation. The most common measure for establishing similarity criteria is Euclidean distance [47].

2.4.4 Naïve bayes (NB)

The NB classifier is frequently utilized in the majority of medical fields for symptom diagnosis [48]. It is supervised learning algorithm based on Bayes theorem. It assumes that the features are statistically independent [45]. The NB is one of the simplest machine learning classifiers for implementation [48]. It determines the likelihood of a particular outcome based on the available dataset. This classifier is considered to be less precise than other methods since it relies on a massive amount of records [48].

The effectiveness of the SVM, DT, KNN, and NB classifiers is evaluated using measures such as accuracy, F-score, and precision. The confusion matrix for this binary classification is shown in Fig. 3. Mathematically, accuracy, precision, and F-score are calculated as follows [24]:

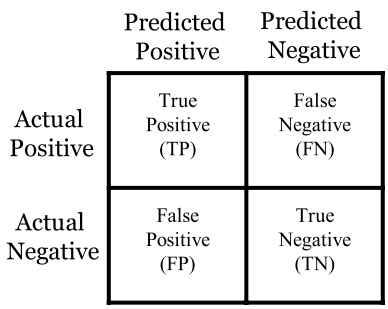

Fig. 3 Confusion matrix of binary classification

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{9}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{10}$$

$$\text{F-score} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}} \tag{11}$$

3 Results & discussion

This section discusses the statistical analysis of proposed TFE features and the performance evaluation of the proposed framework for classifying PD and HC individuals. We have also compared the effectiveness of the proposed TFE features with the breathiness and MFCC features. We have performed statistical analysis using probability density function (PDF) plots to analyze the discriminating potential of the proposed TFE features. The PDF plots for the selected TFE features (TFE_{11}, TFE_{12}, TFE_{15}, TFE_{23}, TFE_{25}, and TFE_{38}) are shown in Fig. 4a–f. It is noted that the peaks of the PDF plots for PD and HC speech classes are located in distinct bins and have different probability values. This shows that TFE features have statistically significant variations; hence we can employ the proposed TFE features for classifying PD and HC individuals. We have used the SVM, DT, KNN, and NB classifiers along with the genetic algorithm for feature selection for classifying PD and HC individuals using TFE, breathiness, and MFCC features. The efficacy of the classifiers for classifying PD ad HC individuals is evaluated using a 10-fold cross-validation.

The performances of the SVM, DT, KNN, and NB classifiers with 10-fold cross-validation for five Spanish sustained vowels and five isolated words are shown in Tables 1 and 2, respectively. All of the results are depicted in terms of *mean* ± *standard deviation*. The classification accuracy achieved with sustained vowels ranges from 72 to 98% using the various machine learning classifiers. The vowel /a/ achieved the highest accuracy of 98% using the SVM classifier. With isolated words, classification accuracy varies from 73 to 99%. The highest classification accuracy of 99% is achieved using the word /atleta/ and the SVM classifier. Conversely, the breathiness and MFCC features perform quite poorly. For the breathiness features, the highest accuracy of 73% is achieved for the word /apto/ using the SVM classifier. The maximum achieved accuracy of MFCC features is 85% for the word /petaka/ using the SVM classifier. The performance of MFCC features is superior to that of breathiness features but inferior to that of CT-based TFE features. In terms of classifier performance, the SVM consistently surpasses the DT, KNN, and NB classifiers.

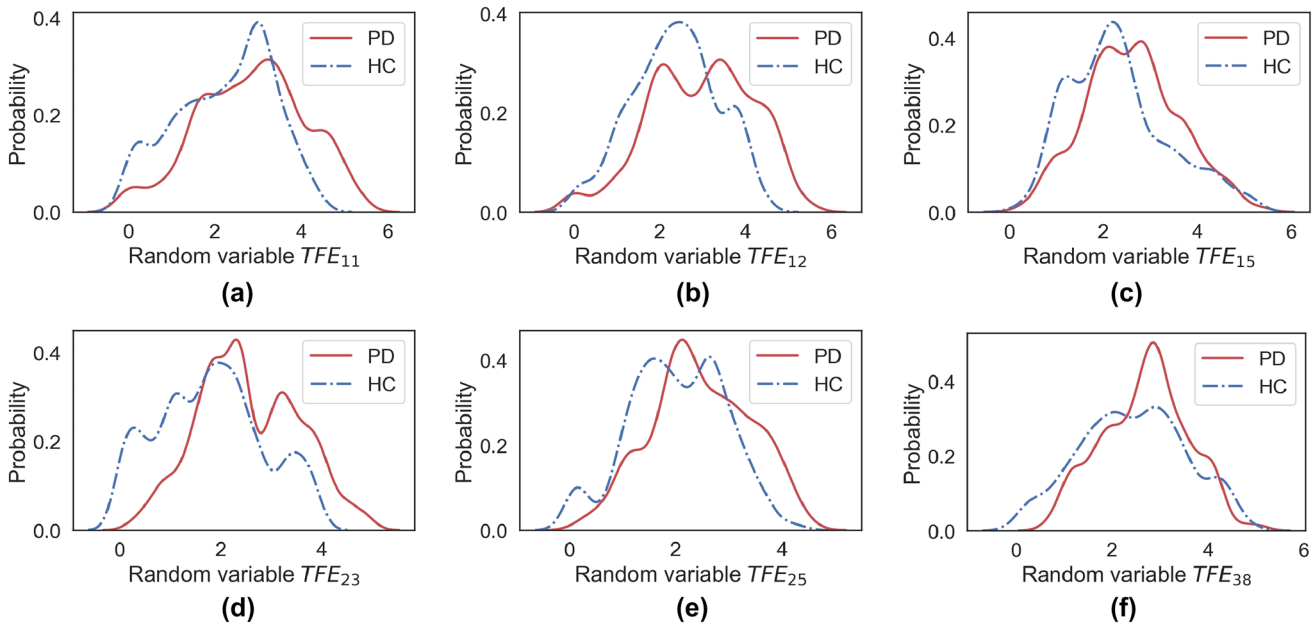

Fig. 4 Probability density function plots of PD and HC speech classes for **a** TFE_{11}, **b** TFE_{12}, **c** TFE_{15}, **d** TFE_{23}, **e** TFE_{25}, and **f** TFE_{38} features

The comparison between the results of the proposed framework and the state-of-the-art methods is presented in Table 3. We have achieved the highest accuracy compared to other state-of-the-art methods. Mehmet et al. [14] enhanced speech recording using variational mode decomposition, and then a spectrogram of enhanced speech signals was used for classifying PD and HC individuals using a combination of CNN and LSTM. Hireš et al. [15] also achieved 99% accuracy on vowel /a/ using an ensemble of CNN classifier, which is a pretty complicated architecture. Conversely, we have achieved the same accuracy using CT-based TFE features and a simple conventional machine learning SVM classifier.

The state-of-the-art methods employed temporal and frequency domain features extensively for diagnosing PD. These commonly used acoustic features may not be efficient in capturing the discontinuities and abrupt changes in speech that occur in PD patients because of the patients weaken muscles and nerves. The proposed CT-based TFE features outperform the conventional MFCC and breathiness features as TFE features jointly capture the temporal and frequency speech perturbation.

The SVM classifier outperforms the other machine learning classifiers, such as DT, KNN, and NB, as well as deep neural network and transformer neural network, because SVM is a highly effective classifier for binary classification that can give excellent performance even with a small learning dataset. The SVM has demonstrated outstanding performance and generalizability when evaluating pathological speech signals [9, 11, 49–53]. The deep learning classifiers and transformer neural networks typically produce superior results with adequate training data. In this study, the dataset size is small; hence their performance lags behind that of the SVM.

Although PD is the second-most common neurodegenerative disease, diagnosing it often takes more than two years. There is no particular diagnostic test for PD. A neurologist diagnoses PD by reviewing the patient medical history, evaluating signs and symptoms, and conducting a physical and neurological examination. We proposed a novel CT-based approach for PD diagnosis utilizing speech signals in this work. First, a TFM representation is obtained for each speech recording using CT, and then TFE features are extracted from it. The genetic algorithm is applied to these features for selecting the efficient subset of features for classifying PD and HC individuals. Results reveal that the proposed TFE features efficiently discriminate between PD and HC speech classes. The discrimination capability of the proposed TFE features is very high compared to the MFCC and breathiness features. The proposed approach uses only speech signals and machine learning techniques for diagnosing PD. Hence, it can be deployed in real-time on an embedded platform to diagnose PD.

4 Conclusion

The diagnosis of PD using speech signals is a breakthrough in the non-invasive diagnosis of various diseases. We presented a novel approach based on the CT in which the TFE

Table 1 Performances of SVM, DT, KNN, and NB classifiers with 10-fold cross-validation for five Spanish sustained vowels

Vowel	Classifier	Proposed TFE features			Breathiness features			MFCC features		
		Accuracy	F-score	Precision	Accuracy	F-score	Precision	Accuracy	F-score	Precision
/a/	SVM	0.980 ± 0.022	0.982 ± 0.021	0.984 ± 0.026	0.690 ± 0.076	0.669 ± 0.096	0.698 ± 0.113	0.837 ± 0.064	0.833 ± 0.065	0.850 ± 0.095
	DT	0.893 ± 0.042	0.889 ± 0.050	0.879 ± 0.053	0.653 ± 0.060	0.647 ± 0.058	0.682 ± 0.161	0.780 ± 0.064	0.780 ± 0.067	0.771 ± 0.092
	KNN	0.900 ± 0.061	0.897 ± 0.062	0.899 ± 0.083	0.660 ± 0.109	0.655 ± 0.124	0.665 ± 0.150	0.800 ± 0.068	0.791 ± 0.062	0.841 ± 0.095
	NB	0.777 ± 0.079	0.756 ± 0.090	0.811 ± 0.100	0.613 ± 0.083	0.589 ± 0.122	0.632 ± 0.184	0.740 ± 0.053	0.739 ± 0.063	0.740 ± 0.108
/e/	SVM	0.970 ± 0.028	0.972 ± 0.024	0.980 ± 0.031	0.693 ± 0.099	0.687 ± 0.104	0.711 ± 0.117	0.840 ± 0.068	0.832 ± 0.089	0.844 ± 0.099
	DT	0.860 ± 0.039	0.859 ± 0.045	0.846 ± 0.080	0.673 ± 0.066	0.683 ± 0.075	0.653 ± 0.094	0.783 ± 0.075	0.780 ± 0.079	0.780 ± 0.099
	KNN	0.873 ± 0.084	0.865 ± 0.080	0.915 ± 0.058	0.637 ± 0.097	0.626 ± 0.100	0.661 ± 0.135	0.800 ± 0.092	0.799 ± 0.082	0.811 ± 0.134
	NB	0.750 ± 0.079	0.748 ± 0.067	0.762 ± 0.091	0.650 ± 0.089	0.611 ± 0.112	0.677 ± 0.130	0.733 ± 0.065	0.740 ± 0.068	0.720 ± 0.114
/i/	SVM	0.970 ± 0.038	0.968 ± 0.039	0.971 ± 0.048	0.707 ± 0.093	0.679 ± 0.083	0.746 ± 0.091	0.823 ± 0.076	0.816 ± 0.092	0.849 ± 0.141
	DT	0.877 ± 0.033	0.876 ± 0.034	0.866 ± 0.065	0.680 ± 0.082	0.670 ± 0.111	0.683 ± 0.152	0.783 ± 0.100	0.784 ± 0.099	0.776 ± 0.110
	KNN	0.880 ± 0.050	0.879 ± 0.041	0.881 ± 0.079	0.690 ± 0.105	0.682 ± 0.114	0.689 ± 0.112	0.843 ± 0.072	0.838 ± 0.081	0.834 ± 0.097
	NB	0.790 ± 0.090	0.788 ± 0.090	0.784 ± 0.096	0.573 ± 0.108	0.353 ± 0.130	0.715 ± 0.166	0.760 ± 0.055	0.746 ± 0.072	0.779 ± 0.088
/o/	SVM	0.933 ± 0.047	0.932 ± 0.052	0.940 ± 0.050	0.693 ± 0.059	0.703 ± 0.067	0.676 ± 0.099	0.787 ± 0.083	0.778 ± 0.088	0.801 ± 0.117
	DT	0.863 ± 0.057	0.861 ± 0.055	0.878 ± 0.086	0.667 ± 0.067	0.662 ± 0.072	0.660 ± 0.089	0.743 ± 0.073	0.734 ± 0.066	0.780 ± 0.105
	KNN	0.883 ± 0.045	0.881 ± 0.043	0.917 ± 0.058	0.643 ± 0.065	0.638 ± 0.068	0.651 ± 0.104	0.787 ± 0.075	0.774 ± 0.089	0.799 ± 0.110
	NB	0.727 ± 0.059	0.732 ± 0.055	0.717 ± 0.047	0.640 ± 0.049	0.667 ± 0.062	0.611 ± 0.073	0.690 ± 0.076	0.693 ± 0.082	0.697 ± 0.147
/u/	SVM	0.930 ± 0.031	0.927 ± 0.035	0.936 ± 0.077	0.707 ± 0.057	0.698 ± 0.084	0.701 ± 0.128	0.830 ± 0.041	0.828 ± 0.042	0.829 ± 0.082
	DT	0.840 ± 0.053	0.832 ± 0.064	0.849 ± 0.067	0.693 ± 0.084	0.699 ± 0.058	0.710 ± 0.131	0.767 ± 0.060	0.762 ± 0.058	0.777 ± 0.065
	KNN	0.853 ± 0.045	0.844 ± 0.057	0.867 ± 0.085	0.663 ± 0.081	0.660 ± 0.104	0.654 ± 0.115	0.823 ± 0.073	0.818 ± 0.077	0.833 ± 0.094
	NB	0.847 ± 0.069	0.840 ± 0.071	0.867 ± 0.088	0.650 ± 0.067	0.638 ± 0.064	0.661 ± 0.117	0.707 ± 0.113	0.687 ± 0.132	0.730 ± 0.149

Table 2 Performances of SVM, DT, KNN, and NB classifiers with 10-fold cross-validation for five isolated words

Word	Classifier	Proposed TFE features			Breathiness features			MFCC features		
		Accuracy	F-score	Precision	Accuracy	F-score	Precision	Accuracy	F-score	Precision
/apto/	SVM	0.950 ± 0.067	0.942 ± 0.075	0.969 ± 0.062	0.730 ± 0.127	0.680 ± 0.166	0.810 ± 0.189	0.810 ± 0.122	0.800 ± 0.128	0.827 ± 0.168
	DT	0.880 ± 0.075	0.872 ± 0.091	0.906 ± 0.133	0.680 ± 0.117	0.639 ± 0.204	0.685 ± 0.217	0.810 ± 0.104	0.802 ± 0.117	0.842 ± 0.202
	KNN	0.850 ± 0.092	0.840 ± 0.081	0.938 ± 0.099	0.680 ± 0.098	0.653 ± 0.147	0.681 ± 0.174	0.800 ± 0.195	0.805 ± 0.152	0.845 ± 0.177
	NB	0.810 ± 0.170	0.790 ± 0.176	0.843 ± 0.173	0.700 ± 0.134	0.620 ± 0.243	0.685 ± 0.271	0.780 ± 0.133	0.775 ± 0.139	0.752 ± 0.185
/atleta/	SVM	0.990 ± 0.030	0.989 ± 0.033	1.000 ± 0.000	0.600 ± 0.126	0.654 ± 0.109	0.598 ± 0.186	0.810 ± 0.104	0.804 ± 0.109	0.833 ± 0.158
	DT	0.900 ± 0.126	0.882 ± 0.157	0.868 ± 0.181	0.670 ± 0.127	0.655 ± 0.149	0.682 ± 0.179	0.830 ± 0.119	0.827 ± 0.136	0.890 ± 0.189
	KNN	0.870 ± 0.119	0.840 ± 0.179	0.930 ± 0.155	0.640 ± 0.150	0.589 ± 0.154	0.672 ± 0.180	0.760 ± 0.092	0.712 ± 0.131	0.820 ± 0.193
	NB	0.830 ± 0.100	0.813 ± 0.116	0.858 ± 0.158	0.600 ± 0.089	0.570 ± 0.141	0.607 ± 0.181	0.780 ± 0.117	0.770 ± 0.134	0.766 ± 0.182
/braso/	SVM	0.880 ± 0.075	0.865 ± 0.084	0.918 ± 0.135	0.690 ± 0.145	0.707 ± 0.148	0.665 ± 0.189	0.800 ± 0.100	0.773 ± 0.116	0.829 ± 0.169
	DT	0.850 ± 0.136	0.842 ± 0.127	0.858 ± 0.163	0.700 ± 0.200	0.690 ± 0.204	0.707 ± 0.199	0.780 ± 0.098	0.770 ± 0.109	0.761 ± 0.133
	KNN	0.790 ± 0.094	0.716 ± 0.137	0.967 ± 0.100	0.640 ± 0.185	0.616 ± 0.233	0.602 ± 0.192	0.700 ± 0.161	0.639 ± 0.264	0.651 ± 0.285
	NB	0.780 ± 0.098	0.746 ± 0.144	0.833 ± 0.216	0.700 ± 0.141	0.686 ± 0.156	0.698 ± 0.180	0.800 ± 0.134	0.812 ± 0.095	0.807 ± 0.121
/globo/	SVM	0.890 ± 0.114	0.883 ± 0.113	0.896 ± 0.111	0.690 ± 0.122	0.654 ± 0.172	0.722 ± 0.201	0.790 ± 0.122	0.756 ± 0.159	0.834 ± 0.163
	DT	0.860 ± 0.080	0.854 ± 0.091	0.832 ± 0.145	0.650 ± 0.120	0.621 ± 0.167	0.645 ± 0.153	0.830 ± 0.149	0.795 ± 0.180	0.857 ± 0.198
	KNN	0.770 ± 0.142	0.733 ± 0.149	0.861 ± 0.173	0.680 ± 0.140	0.653 ± 0.191	0.654 ± 0.179	0.690 ± 0.176	0.628 ± 0.275	0.675 ± 0.326
	NB	0.730 ± 0.127	0.735 ± 0.126	0.695 ± 0.156	0.600 ± 0.110	0.672 ± 0.094	0.604 ± 0.186	0.750 ± 0.136	0.775 ± 0.112	0.765 ± 0.205
/petaka/	SVM	0.940 ± 0.066	0.936 ± 0.071	0.912 ± 0.119	0.670 ± 0.110	0.605 ± 0.155	0.753 ± 0.186	0.850 ± 0.102	0.861 ± 0.074	0.852 ± 0.141
	DT	0.920 ± 0.087	0.888 ± 0.126	0.863 ± 0.167	0.630 ± 0.142	0.631 ± 0.156	0.638 ± 0.203	0.810 ± 0.104	0.810 ± 0.086	0.860 ± 0.196
	KNN	0.860 ± 0.092	0.853 ± 0.085	0.947 ± 0.111	0.700 ± 0.110	0.637 ± 0.180	0.795 ± 0.187	0.760 ± 0.150	0.725 ± 0.154	0.848 ± 0.160
	NB	0.850 ± 0.175	0.863 ± 0.143	0.913 ± 0.108	0.640 ± 0.111	0.590 ± 0.152	0.672 ± 0.166	0.790 ± 0.122	0.774 ± 0.132	0.798 ± 0.171

Table 3 Performance comparison with the state-of-the-art methods on the PC-GITA database

Research work	Feature	Classifier	Achieved accuracy
Orozco-Arroyave et al. [9]	Spectral and cepstral features	SVM	76% for vowel /e/ 89% for word /petaka/
Karan et al. [10]	IEDCC	SVM	90%for vowel /o/ 91% for word /globo/
Karan et al. [11]	Time-frequency features	SVM	92% for vowel /e/ 97% for word /petaka/
Narendra et al. [12]	End-to-end approach using QCP-based glottal signal	CNN + LSTM	68.56% for continuous speech
Karan et al. [13]	HCC	MLP	91% for vowel /a/ 96% for word /apto/
Mehmet et al. [14]	Spectrogram	ResNet + LSTM	98.61% for words and vowels
Hireš et al. [15]	Spectrogram	CNN	99% for vowel /a/
Quan et al. [16]	Time series dynamic features using time-distributed 2D-CNN	CNN	85.33% for vowel /a/ 81.50% for word /apto/
Proposed method	CT based TFE features	SVM	98% for vowel /a/ 99% for word /atleta/

features are extracted from the TFM representation of the speech obtained using the CT. The efficacy of the proposed method is accessed using the words and sustained vowels from the PC-GITA database. The proposed TFE features outperform the breathiness and MFCC features. The highest results of 98% and 99% are achieved with sustained vowel /a/ and word /atleta/, respectively. The results demonstrate the usefulness of the presented CT-based framework for diagnosing PD using speech signals. We can conclude that the CT-based TFM representation of speech shows significant variations for PD and HC individuals. These variations can be captured using the TFE features for classifying PD and HC individuals.

In further work, we will examine the competency of the developed framework on a large dataset to better comprehend its competence. Additionally, we'll use the developed framework to assess the degree of dysarthria of PD patients.

Author Contributions All authors made intellectual contributions to the study. PW and SPM conducted the research. PW, SPM, and SD prepared the manuscript. SD reviewed the manuscript. All of the authors have read and approved the manuscript.

Funding Not applicable.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships.

Ethical approval Not applicable.

References

1. Orozco-Arroyave JR, Arias-Londoño JD, Vargas-Bonilla JF, Gonzalez-Rátiva MC, Nöth E. New spanish speech corpus database for the analysis of people suffering from parkinson's disease. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), 2014;p. 342–347.
2. De Rijk Md, Launer L, Berger K, Breteler M, Dartigues J, Baldereschi M, Fratiglioni L, Lobo A, Martinez-Lage J, Trenkwalder C, et al. Prevalence of parkinson's disease in europe: A collaborative study of population-based cohorts. Neurologic diseases in the elderly research group. Neurology. 2000;54(11):S21-3.
3. Rana AQ, Ahmed US, Chaudry ZM, Vasan S. Parkinson's disease: a review of non-motor symptoms. Expert Rev Neurother. 2015;15(5):549–62.
4. Hughes AJ, Daniel SE, Ben-Shlomo Y, Lees AJ. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain. 2002;125(4):861–70.
5. Ramig LO, Fox C, Sapir S. Speech treatment for parkinson's disease. Expert Rev Neurother. 2008;8(2):297–309.
6. Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, Apaydin H, Kursun O. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inform. 2013;17(4):828–34.
7. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease. IEEE Trans Biomed Eng. 2012;59(5):1264–71.
8. Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L, Farahmand M. A hybrid intelligent system for the prediction of parkinson's disease progression using machine learning techniques. Biocybern Biomed Eng. 2018;38(1):1–15.
9. Orozco-Arroyave JR, Hönig F, Arias-Londoño JD, Vargas-Bonilla JF, Nöth E. Spectral and cepstral analyses for Parkinson's disease detection in Spanish vowels and words. Expert Syst. 2015;32(6):688–97.
10. Karan B, Sahu SS, Orozco-Arroyave JR, Mahto K. Hilbert spectrum analysis for automatic detection and evaluation of parkinson's speech. Biomed Signal Process Control. 2020;61: 102050.
11. Karan B, Sahu SS, Orozco-Arroyave JR, Mahto K. Non-negative matrix factorization-based time-frequency feature extraction of voice signal for parkinson's disease prediction. Comput Speech Lang. 2021;69: 101216.
12. Narendra N, Schuller B, Alku P. The detection of Parkinson's disease from speech using voice source information. IEEE/ACM Trans Audio Speech Lang Process. 2021;29:1925–36.
13. Karan B, Sahu SS. An improved framework for Parkinson's disease prediction using variational mode decomposition-Hilbert spectrum of speech signal. Biocybern Biomed Eng. 2021;41(2):717–32.

14. Er MB, Isik E, Isik I. Parkinson's detection based on combined cnn and lstm using enhanced speech signals with variational mode decomposition. Biomed Signal Process Control. 2021;70: 103006.
15. Hireš M, Gazda M, Drotár P, Pah ND, Motin MA, Kumar DK. Convolutional neural network ensemble for Parkinson's disease detection from voice recordings. Comput Biol Med. 2022;141: 105021.
16. Quan C, Ren K, Luo Z, Chen Z, Ling Y. End-to-end deep learning approach for Parkinson's disease detection from speech signals. Biocybern Biomed Eng. 2022;42(2):556–74.
17. Karan B, Sahu SS, Mahto K. Parkinson disease prediction using intrinsic mode function based features from speech signal. Biocybern Biomed Eng. 2020;40(1):249–64.
18. Fujita T, Luo Z, Quan C, Mori K, Cao S. Performance evaluation of rnn with hyperbolic secant in gate structure through application of Parkinson's disease detection. Appl Sci. 2021;11(10):4361.
19. Goyal J, Khandnor P, Aseri TC. A hybrid approach for Parkinson's disease diagnosis with resonance and time-frequency based features from speech signals. Expert Syst Appl. 2021;182: 115283.
20. Mann S, Haykin S. The Chirplet transform: physical considerations. IEEE Trans Signal Process. 1995;43(11):2745–61.
21. Peng Z, Meng G, Chu F, Lang Z-Q, Zhang W, Yang Y. Polynomial Chirplet transform with application to instantaneous frequency estimation. IEEE Trans Instrum Meas. 2011;60(9):3222–9.
22. Rabiner L, Juang B-H. Fundamentals of speech recognition, Prentice-Hall, Inc., 1993.
23. Tripathy RK, Paternina MR, Arrieta JG, Zamora-Méndez A, Naik GR. Automated detection of congestive heart failure from electrocardiogram signal using stockwell transform and hybrid classification scheme. Comput Methods Programs Biomed. 2019;173:53–65.
24. Ghosh SK, Ponnalagu R, Tripathy R, Acharya UR. Automated detection of heart valve diseases using Chirplet transform and multiclass composite classifier with pcg signals. Comput Biol Med. 2020;118: 103632.
25. Castillo-Guerra E, Ruiz A. Automatic modeling of acoustic perception of breathiness in pathological voices. IEEE Trans Biomed Eng. 2008;56(4):932–40.
26. Deb S, Dandapat S. A novel breathiness feature for analysis and classification of speech under stress. In: 2015 Twenty First National Conference on Communications (NCC), IEEE, 2015;p. 1–5.
27. Chandrashekar H, Karjigi V, Sreedevi N. Breathiness indices for classification of dysarthria based on type and speech intelligibility. In: 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), IEEE, 2019;p. 266–270.
28. Deb S, Warule P, Nair A, Sultan H, Dash R, Krajewski J. Detection of common cold from speech signals using deep neural network, Circuits, Systems, and Signal Processing. 2022;p. 1–16.
29. Warule P, Mishra SP, Deb S. Classification of cold and non-cold speech using vowel-like region segments. In: 2022 IEEE International Conference on Signal Processing and Communications (SPCOM), IEEE, 2022;p. 1–5.
30. Warule P, Mishra SP, Deb S, Krajewski J. Sinusoidal model-based diagnosis of the common cold from the speech signal. Biomed Signal Process Control. 2023;83: 104653.
31. Goldberg DE, Genetic algorithms, Pearson Education India, 2013.
32. Kinnear KE, Langdon WB, Spector L, Angeline PJ, O'Reilly U-M. Advances in genetic programming, vol. 3, MIT press, 1994.
33. Murthy YS, Koolagudi SG. Classification of vocal and non-vocal segments in audio clips using genetic algorithm based feature selection. Expert Syst Appl. 2018;106:77–91.
34. Hosny M, Zhu M, Gao W, Fu Y. A novel deep learning model for stn localization from lfps in Parkinson's disease. Biomed Signal Process Control. 2022;77: 103830.
35. Schuller B, Rigoll G, Lang M. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In: 2004 IEEE international conference on acoustics, speech, and signal processing, Vol. 1, IEEE, 2004;p. I–577.
36. Shahbakhi M, Far DT, Tahami E. Speech analysis for diagnosis of parkinson's disease using genetic algorithm and support vector machine. J Biomed Sci Eng. 2014;2014:
37. Deb S, Dandapat S, Krajewski J. Analysis and classification of cold speech using variational mode decomposition. IEEE Trans Affect Comput. 2017;11(2):296–307.
38. Warule P, Mishra SP, Deb S. Significance of voiced and unvoiced speech segments for the detection of common cold. Signal Image Video Process 2022;1–8.
39. Pishgar M, Karim F, Majumdar S, Darabi H. Pathological voice classification using mel-cepstrum vectors and support vector machine. (2018). arXiv preprint arXiv:1812.07729.
40. Noble WS. What is a support vector machine? Nature Biotechnol. 2006;24(12):1565–7.
41. Ghosh SK, Tripathy RK, Ponnalagu R. Classification of pcg signals using fourier-based synchrosqueezing transform and support vector machine. IEEE Sensors. 2021;2021:1–4.
42. Ghiasi MM, Zendehboudi S. Decision tree-based methodology to select a proper approach for wart treatment. Comput Biol Med. 2019;108:400–9.
43. Song Y-Y, Ying L. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry. 2015;27(2):130.
44. Hussein RM, George LE, Miften FS. Accurate method for sleep stages classification using discriminated features and single eeg channel. Biomed Signal Process Control. 2023;84: 104688.
45. Patgiri C, Ganguly A. Adaptive thresholding technique based classification of red blood cell and sickle cell using Naïve bayes classifier and k-nearest neighbor classifier. Biomed Signal Process Control. 2021;68: 102745.
46. Mazzoleni M, Previdi F, Bonfiglio NS. Classification algorithms analysis for brain-computer interface in drug craving therapy. Biomed Signal Process Control. 2019;52:463–72.
47. Acosta-Mesa H-G, Cruz-Ramírez N, Hernández-Jiménez R. Aceto-white temporal pattern classification using k-nn to identify precancerous cervical lesion in colposcopic images. Comput Biol Med. 2009;39(9):778–84.
48. Olatunji SO, Alotaibi S, Almutairi E, Alrabae Z, Almajid Y, Altabee R, Altassan M, Ahmed MIB, Farooqui M, Alhiyafi J. Early diagnosis of thyroid cancer diseases using computational intelligence techniques: a case study of a Saudi Arabian dataset. Comput Biol Med. 2021;131: 104267.
49. Rusz J, Cmejla R, Ruzickova H, Ruzicka E. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. J Acoust Soc Am. 2011;129(1):350–67.
50. Bocklet T, Nöth E, Stemmer G, Ruzickova H, Rusz J. Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis. IEEE Workshop Autom Speech Recogn Underst. 2011;2011:478–83.
51. Vásquez-Correa JC, Orozco-Arroyave J, Bocklet T, Nöth E. Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease. J Commun Disord. 2018;76:21–36.

52. Orozco-Arroyave JR, Belalcazar-Bolanos EA, Arias-Londoño JD, Vargas-Bonilla JF, Skodda S, Rusz J, Daqrouq K, Hönig F, Nöth E. Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases. IEEE J Biomed Health Inform. 2015;19(6):1820–8.
53. Maier A, Haderlein T, Eysholdt U, Rosanowski F, Batliner A, Schuster M, Nöth E. Peaks-a system for the automatic evaluation of voice and speech disorders. Speech Commun. 2009;51(5):425–37.