Available online at www.sciencedirect.com

ScienceDirect

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

Original Research Article

Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features

Gabriel Solana-Lavalle, Juan-Carlos Galán-Hernández, Roberto Rosas-Romero *

Universidad de las Américas-Puebla, Puebla, Mexico

ARTICLE INFO

Article history:
Received 8 October 2019
Received in revised form
3 December 2019
Accepted 1 January 2020
Available online 05 February 2020

Keywords:
Parkinson disease
Dysphonia features
Feature subset selection
Support vector machine

ABSTRACT

Recent research on *Parkinson disease* (PD) detection has shown that vocal disorders are linked to symptoms in 90% of the PD patients at early stages. Thus, there is an interest in applying vocal features to the computer-assisted diagnosis and remote monitoring of patients with PD at early stages. The contribution of this research is an increase of accuracy and a reduction of the number of selected vocal features in PD detection while using the newest and largest public dataset available. Whereas the number of features in this public dataset is 754, the number of selected features for classification ranges from 8 to 20 after using *Wrappers feature subset selection*. Four classifiers (*k nearest neighbor*, *multi-layer perceptron*, *support vector machine* and *random forest*) are applied to vocal-based PD detection. The proposed approach shows an *accuracy* of 94.7%, *sensitivity* of 98.4%, *specificity* of 92.68% and *precision* of 97.22%. The best resulting accuracy is obtained by using a *support vector machine* and it is higher than the one, which was reported on the first work to use the same dataset. In addition, the corresponding computational complexity is further reduced by selecting no more than 20 features.

1. Introduction

Parkinson disease (PD) is a chronic and degenerative neurological disease that affects 1% of people over age 60 and it is the most common neurological disorder after Alzheimer. PD is the result of losing cells in different regions of the brain, mainly the *substantia nigra*. The substantia nigra produces *dopamine*, a carrier of the signals for movement coordination. The loss of dopamine causes neurons to fire without control so that these patients lose the ability to control limb movement [1].

The cardinal symptoms of PD patients are *motor symptoms* such as a resting tremor where limbs move without control (diskinesya), slowness of movement (bradykinesia), postural instability (balance problems) and rigidity [1]. Since motor symptoms are revealed at an advanced disease stage, PD is difficult to diagnose at an early stage. Although PD is not cured, the quality of life in PD patients is improved when they are

* *Corresponding author at*: Universidad de las Américas-Puebla, Mexico.
E-mail address: roberto.rosas@udlap.mx (R. Rosas-Romero).
https://doi.org/10.1016/j.bbe.2020.01.003

treated promptly as a result of early diagnosis [2]. On the other hand, there are symptoms, which are *non-motor* symptoms such as cognitive impairment, mood disorders, sleep difficulties, loss of sense of smell, speech and swallowing problems, deterioration of writing skills, drooling, and low blood pressure when standing [1].

The vocal disorders are critical and *prodromal* (brain phase before altered brain stages) symptoms since they manifest in 90% of the PD patients at an early stage [3]. Vocal disorders in PD patients are dysphonia (defective voice), hypophonia (reduced volume), monotone (reduced pitch), and dysarthia (difficulty with articulation) [3]. Vocal feature extraction has been applied to PD detection at early stages [3–9]. The monitoring of PD evolution has also been possible by analyzing speech [10,6]. Furthermore, speech-based telediagnosis and telemonitoring are non-invasive, simple to use and low in cost [11,12].

The most commonly used vocal features, known as *baseline features*, include *jitter*, *shimmer*, *fundamental frequency parameters*, *harmonicity parameters*, *recurrence period density entropy* (RPDE), *detrended fluctuation analysis* (DFA) and *pitch period entropy* (PPE) [13]. The analysis of *dysphonia* (inability to produce normal vocal sounds), through signal processing techniques, has also been applied to PD detection [10,12,14], where an accuracy of 90% has been reached [10]. One of the most comprehensive analysis of dysphonia features for PD detection, is based on 132 features grouped into three main subsets [6]. Furthermore, different vocal feature subsets have been merged into one single set, on which *feature subset selection* is subsequently applied to reduce the number of features by choosing the most relevant [6].

Research, oriented to vocal-based PD detection, has been possible because of the availability of public datasets; however, these datasets have the disadvantage that vocal feature extraction was performed on a small population of PD patients and healthy subjects. In this work, vocal-based PD detection is approached by (1) using the newest and largest public dataset (64 healthy individuals and 188 PD patients), (2) selecting a small set of features and (3) increasing detection accuracy. The used dataset was collected at the Istanbul University [13]. In this work, PD detection, based on the largest dataset, is improved in terms of the classification accuracy and the number of selected features.

Three sets of experiments were conducted in the work, in which the largest available dataset was introduced [13]. First, accuracy measures were obtained by exhaustively combining six different feature subsets (baseline, wavelet-based, Mel-frequency cepstral coefficients – MFCC, intensity-based, vocal fold-based, bandwidth) and six different classifiers (Naive Bayes, logistic regression, k Nearest Neighbor – *kNN*, multilayer perceptron – *MLP*, random forest, support vector machine – *SVM*). In the first set of experiments, the highest achieved accuracy was 0.84, where the corresponding feature subset was the MFCC set and the corresponding classifier was a SVM with a radial basis function (*RBF*) kernel. In a second set of experiments, the *tunable Q-factor wavelet transform* (TQWT) was used for vocal feature extraction for the first time [15,16], with the result that the highest reached accuracy was 0.85 by using a MLP. In a third set of experiments, the 50 most relevant features were obtained by applying *minimum-redundancy maxim-relevance* (*mRMR*) feature selection to the collection of all feature subsets [17], which jointly contain 754 features. The highest accuracy achieved with the top-50 features was 0.86 based on a SVM-RBF classifier.

The contribution of our work consists in improvements over the work of Sakar et al. [13], in terms of complexity and accuracy. First, the number of the selected top features is reduced from 50 to 8–20 by applying *Wrappers feature subset selection* [18] to the whole set of 754 features extracted from 252 patients in the largest dataset for voice-based PD detection. Second, the current proposed method reaches an accuracy of 0.94, which is higher than the previous highest accuracy (0.86). The proposed methodology is based on the use of four classifiers (kNN, MLP, SVM-RBF, Random Forest – RF), Wrappers feature subset selection, the largest dataset for PD detection and a small collection of the most relevant and uncorrelated features.

Section 2 discusses how vocal features, feature subset selection, classification and validation are applied to the task of PD detection. Section 3 provides the experimental results obtained by following different strategies of feature subset selection and classification as well as comparisons. The results are discussed in Section 4. The conclusions are presented in Section 5.

2. Method

The proposed method for vocal-based PD detection consists of different stages. First, speech recordings were obtained from multiple individuals, and *feature extraction* was performed by applying different signal processing techniques. The signal recording and feature extraction were performed at the Istambul University [13]. In a second stage, all features were standardized to have zero mean and unit standard deviation since each feature has a different range of values. The aim of the subsequent stage, *feature subset selection*, is to find a reduced number of features, which are relevant and uncorrelated. The last stages consist in classifying observed features and assessing detection performance. An overview of the proposed method is shown in Fig. 1. The red rectangle encloses the stages carried out for this work.

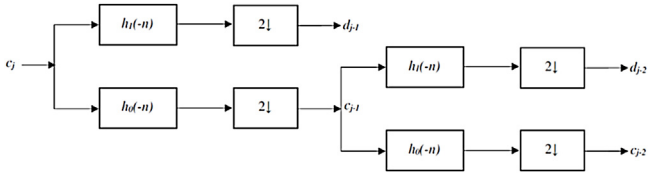

Fig. 1 – Overview of the proposed approach for PD detection.

2.1. Dataset description

The dataset for this study is the *Parkinson's Disease Classification dataset*, which is found within the *Machine Learning Repository* of the University of California Irvine. This dataset was generated by the *Cerrahpsa Faculty of Medicine* at the *Department of Neurology*, *Istanbul University*, from 188 PD patients (107 men and 81 women) with ages ranging from 33 to 87 years old (65.1 years old $\pm$ 10.9 years), and from 64 healthy individuals (23 men and 41 women) with ages from 41 to 82 years old (61.1 years old $\pm$ 8.9 years). According to the Braak model [19], as the Parkinson disease evolves, the autonomic, limbic, and somatomotor systems become damaged. During the pre-symptomatic stages 1 and 2 of the disease (45 years old $\pm$ 5 years), the pathology is confined to the medulla oblongata and the olfactory bulb. The medulla oblongata is responsible for involuntary functions and it is associated to those innervated muscles such as the tongue, pharynx and larynx (speech problems). In the stages 3 and 4 (55 years old $\pm$ 5 years), one of the pathological changes is the loss of dopamine, which plays an important role in motor functions such as speech. In the final stages 5 and 6, the disease manifests in all of its main symptoms. Vocal features are used to assist in the detection of PD at early stages since speech problems are revealed during stages 1–4 of the Parkinson disease.

The voice samples were recorded after the subjects under study were diagnosed as being healthy or having Parkinson disease. All subjects were informed about the dataset generation process, signed informed consent, and attended the test voluntarily in accordance with the approval of Clinical Research Ethics Committee of Bahcesehir University. The sustained phonation of the vowel /a/ is recorded from each of 252 individuals during three repetitions (756 instances), where the total number of speech instances recorded from healthy subjects is 193 while the corresponding number for PD patients is 564. Statistical analysis has been performed to compare each acoustic parameter (jitter, shimmer, noise-to-harmonics-ratio, and F_1) of the vowel sounds between patients with PD and normal controls, with the general result that values of parameters for vowel /a/ in PD patients show a significant difference when they are compared to those of the normal controls [20]. This is the reason for using vowel /a/.

A feature vector contains 754 clinically useful attributes, which were extracted from a recording instance by using different speech processing techniques [13] within the Praat acoustic analysis software [21].

2.2. Feature selection

A *feature subset selection* problem consists in finding the most useful subset of features suited to the used classifier, allowing the classifier to obtain better results and reduce the complexity of the model. For the case of n features, the feature selection process results in m features such that $m < n$, where m is the smallest set of *significant* and *relevant* features. The motivations for feature subset selections is that classification accuracy is diminished when many unnecessary features are used and/or features are correlated [18]. There are different feature subset selection algorithms, such as *stepwise selection* and *minimum-redundancy-maximum-relevance* (*mRMR*) feature selection [22].

While the best subset selection algorithms analyzes all $2^n = 2^{754}$ possible subsets of the $n = 754$ features, the *Wrappers feature subset selection* is a computationally efficient alternative because it considers a reduced number of these subsets. It combines *forward stepwise selection* and *backward stepwise selection*. *Forward stepwise selection* begins with a model with no features, and it adds features to the model, one at a time. At each step, the feature, that gives the best improvement, is added to the model. *Backward stepwise selection* begins with all n features, and it iteratively removes the least useful feature, one at a time. *Wrappers* adds a new feature while it also checks the relevance of already added features. If it finds an insignificant feature, then it removes that particular feature. The steps involved are:

1. Let M_0 be an empty model, which contains no features.
2. Choose a significance level to *add* L_{add} or *drop* L_{drop} a feature from the model.
3. For $k = 0, 1, \ldots, n-1$: (a) Analyze all $n-k$ models that add one additional feature to M_k. (b) Among the $n-k$ models, the best model is selected as M_{k+1} by using the *cross validated error* E so that a newly added feature must satisfy $L_{add} < E$ to be added. (c) Perform backward elimination of any previously added feature with $L_{drop} > E$.
4. Repeat step 3 until a final optimal set of features is obtained among $M_1, M_2, \ldots, M_n$.

The error function, used to determine the best model, is computed according to $E = \sum_i (y_i - d_i)^2$, where y_i is the classifier output associated to feature vector $\mathbf{x}_i$, and d_i is the corresponding desired value.

2.3. Features

During generation of the dataset, the sustained phonation of vowel /a/ was recorded from each individual. The recording time for each recorded instance was 220 s through the use of a microphone with a sampling frequency of 44 kHz. Because speech signals are non-stationary, they are divided into frames. Each frame consists of 1100 samples, corresponding to 0.025 s time duration, where the signal is assumed stationary, and with an overlap percentage between two adjacent frames of 275 samples (25%). A total of 754 vocal features were extracted from each recorded instance [13]. Since each instance is divided into frames, features are initially extracted per frame, and then statistical features (*mean value*, *standard deviation* – std, *skewness*, *kurtosis*, *minimum value* – min, *entropy*) are computed to characterize a complete instance (voice recording).

In the proposed method, few vocal features are used after selecting the least redundant and most relevant features. The selected features correspond to four sets of vocal features: *baseline features*, *Mel frequency cepstral coefficients* - MFCC, *wavelet features* – WT, and *tunable Q-factor wavelet transform* – TQWT. This subsection focuses on a general overview of the framework for the generation of the used vocal features.

2.3.1. Baseline features

The *baseline feature set* has been one of the most commonly used sets in speech analysis. For the case of the proposed

method, only two baseline features were used, *jitter* and *detrended fluctuation analysis* (DFA).

Jitter is a measure of the frequency instability within the oscillating patterns of a vocal time series. *Jitter* is the average absolute difference of the cycle-to-cycle fundamental period, $jitter = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}|$, where T_i is the extracted period at cycle i and N is the number of analyzed cycles.

The second baseline feature, *DFA*, detects long-range correlations in multi-segment time series [23,24]. The computation of the DFA for a time series $\{x(n); n = 1, \ldots, N\}$ consists of different steps: (1) An accumulated version of the zero-mean time series is computed, $y(m) = \sum_{n=1}^{m}[x(n) - x]$. (2) $y(m)$ is partitioned into K non-overlapping segments $\{S_i; i = 1, \ldots, K\}$ of equal length $s = \frac{N}{K}$. (3) At each segment S_i, a straight line $y_{LS}(m)$ is fit to $y(m)$ by using Least-Squares linear regression. (4) The root-mean-square fluctuation of an accumulated and detrended time series is computed:

$$F(s) = \sqrt{\frac{1}{N}\sum_{m=1}^{N}\left[y(m) - y_{LS}(m)\right]^2}. \tag{1}$$

This function $F(s)$ is evaluated over different segment size values s. (5) A power law relation between $F(s)$ and s is obtained, $F(s) \propto s^H$, where the parameter H (*Hurst exponent*), represents the correlation properties of the time series. The scaling exponent H is computed as the slope of a straight line fit to the log-log curve of $F(s)$ vs. s using least-squares linear regression. If $H = 0.5$ then $x(n)$ is uncorrelated. If $H < 0.5$ then $x(n)$ is anticorrelated. If $H > 0.5$, there are positive correlations in $x(n)$.

2.3.2. *Mel frequency cepstral coefficients*

The *Mel Frequency Cepstral Coefficients* (MFCC) have been applied to speech recognition and have been useful for detection of motion changes in tongue and lips, which are affected by PD [4,5]. The computation of the MFCCs consists of different steps [25,26]: (1) The phonation time series $x(n)$ is divided into N-sample segments $\{x(n)w(n);\ n = 1, \ldots, N\}$, where $w(n)$ is a *Hamming* window. (2) The *discrete Fourier transform* (DFT) of each segment is computed:

$$X(k) = \sum_{n=0}^{N-1} w(n)x(n)e^{-j\frac{2\pi}{N}kn}; k = 0, \ldots, N-1. \tag{2}$$

(3) Voice signals do not follow a linear frequency scale and that is why the magnitude spectrum $|X(k)|$ is scaled in frequency and magnitude. The mapping of the acoustic frequency scale is known as the *Mel scale*, $f_{mel} = 2595\log_{10}(1 + \frac{f}{700})$ [27]. (4) The magnitude spectrum $|X(k)|$ is processed by a *Mel filter bank* $\{H(k, m); m = 1, \ldots, M\}$, where M is the number of filters. At each band, the logarithm of the weighted sum of the DFT coefficients $|X(k)|$ is computed:

$$Y(m) = \ln\sum_{k=0}^{N-1}|X(k)|H(k, m); m = 1, \ldots, M. \tag{3}$$

The sequence $Y(m)$ is zero everywhere except at the centers of the frequency bands $\{f_c(m); m = 1, \ldots, M\}$. The Mel filter bank is a set of twelve triangular band-pass filters uniformly distributed on the Mel scale, with central frequencies $\{f_c(m); m = 1, \ldots, M\}$, and it is defined as:

$$H(k, m) = \begin{cases} 0 & \text{for } f(k) < f_c(m-1), \\ \dfrac{f(k) - f_c(m-1)}{f_c(m) - f_c(m-1)} & \text{for } f_c(m-1) \le f(k) \le f_c(m), \\ \dfrac{f(k) - f_c(m+1)}{f_c(m) - f_c(m+1)} & \text{for } f_c(m) \le f(k) \le f_c(m+1), \\ 0 & \text{for } f(k) \ge f_c(m+1). \end{cases} \tag{4}$$

(4) The *Mel frequency cepstral coefficients* (MFCC) are obtained by applying the *discrete cosine transform* (DCT) [27],

$$c_{mel}(n) = \sum_{m=0}^{M-1} Y(m)\cos\left(\frac{\pi}{M}\left(m - \frac{1}{2}\right)n\right); n = 0, 1, \ldots, N-1. \tag{5}$$

There are features related to the change in the cepstral coefficients over time, which are called *delta coefficients*. Typical MFCC-related features are the 12 MFCC coefficients (the mean of the sixth coefficient was selected by Wrappers for the case of the SVM classifier), 12 delta (or velocity) MFCC features (the mean of the third delta was selected by Wrappers for the MLP, the standard deviations of the seventh delta and ninth delta were selected by Wrappers for the SVM), 12 double-delta (or acceleration) MFCC features (the standard deviations of the first and ninth double-deltas were selected by Wrappers for the RF classifier), 1 energy feature,1 delta energy feature, 1 double-delta energy feature (selected by Wrappers for the case of the kNN, MLP and SVM). The *delta coefficients* are computed according to $d(t) = \frac{\sum_{n=1}^{N} n(c(t+n) - c(t-n))}{2\sum_{n=1}^{N} n^2}$, where a typical value for N is 2. Double-deltas (or acceleration) coefficients are calculated in the same way from the deltas.

2.3.3. *Discrete wavelet transform features*

The *discrete wavelet transform* (DWT) for a time series $x(n)$ is expressed as a linear combination of time-shifted versions of a *scaling function* $\phi_{j,k}(n) = \phi(2^j t - k)$ and time-shifted and time-dilated versions of a *mother wavelet* function $\psi_{j,k}(n) = \psi(2^j t - k)$

$$x(n) = \sum_{k=-\infty}^{\infty} c_k\phi_{j_0,k} + \sum_{j=j_0}^{\infty}\sum_{k=-\infty}^{\infty} d_{j,k}\psi_{j,k}, \tag{6}$$

where k corresponds to a *time shift*, j corresponds to a time dilation, $\{d_{j,k}\}$ are the *wavelet coefficients*, and $\{c_k\}$ are the *scaling coefficients* [28]. The coefficients, c_{j-1} and d_{j-1}, are iteratively obtained by filtering the coefficients c_j at a higher resolution scale with a low-pass filter $h_0(-n)$ and a high-pass filter $h_1(-n)$, as it is shown in Fig. 2.

The first stage of the *Wavelet filter bank* divides the spectrum of $x(n) = c_j$ into two equal parts, a low-pass band and a high-pass band corresponding to c_{j-1} and d_{j-1}, respectively. The subsequent stage divides the lower spectrum half c_{j-1} into two equal quarters corresponding to c_{j-2} and d_{j-2}, respectively. In these filters, the ratio of the central frequency over the bandwidth is constant, $Q = \frac{\omega_0}{B} = 1.5$.

2.3.4. *Tunable Q-factor wavelet transform features*

The *tunable Q-factor wavelet transform* (TQWT) is implemented by iteratively filtering the lower part of a two-band signal with

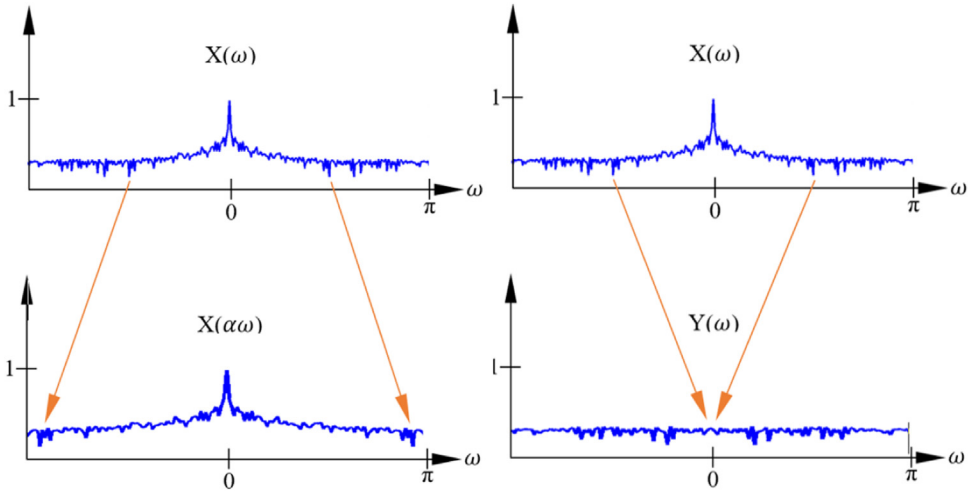

Fig. 2 – Two-stage filter bank.

a low-pass scaling filter and a high-pass scaling filter [15]. Low-pass scaling, with a factor $0<\alpha<1$, generates a signal with a sampling rate of αf_s, where f_s is the sampling rate of the input signal $x(n)$. Low-pass scaling preserves the low-frequency content of $x(n)$, as it is shown in the left side of Fig. 3, and it is defined as:

$$Y(\omega) = X(\alpha\omega), |\omega| < \pi. \tag{7}$$

On the other hand, high-pass scaling preserves the high-frequency content. When $0<\beta<1$, high-pass scaling is defined as:

$$Y(\omega) = \begin{cases} X(\beta\omega + (1-\beta)\pi) & \text{for } 0<\omega<\pi, \\ X(\beta\omega - (1-\beta)\pi) & \text{for } -\pi<\omega<0. \end{cases} \tag{8}$$

The TQWT is based on the filter bank of Fig. 4. The low-pass signal $\overset{(1)}{0}(n)$ and high-pass signal $\overset{(1)}{1}(n)$, at the first stage, have sampling rates αf_s and βf_s, respectively. $H_0^{(j)}(\omega)$ and $H_1^{(j)}(\omega)$ are the low-pass and high-pass frequency responses at the jth stage. By varying α and β, the frequency decomposition is adjusted. The jth stage Q factor is expressed in terms of β as $Q = \frac{\omega c}{BW} = \frac{2-\beta}{\beta}$.

2.4. *Classifiers*

In this work, we are using four classification techniques; kNN, SVM, MLP and Random Forest.

The *k Nearest Neighbor* (kNN) is a non-linear classifier. Within N stored feature vectors $\{x_i; i = 1,\ldots,N\}$, k feature vectors are identified as the *nearest neighbors* of the unknown x, according to $x_i = \text{argmin}||x - x_i||^2$. Each neighbor x_i belongs to one class, and there are two classes, healthy and PD patient. The number of nearest neighbors k_i from class C_i $(i = 1, 2)$ is determined, where $k = k_1 + k_2$. The class, assigned to x, is the one with the largest k_i. For the implementation of the kNN classifier, k = 3.

The *multi-layer perceptron* (MLP) is an artificial neural network (ANN) with multiple layers of neurons. A neuron at any layer feeds all neurons from a subsequent layer, according to $y_j^\ell = \sum_{i=1}^{n_{\ell-1}} \omega_{j,i}^\ell y_i^{\ell-1}$; where y_j^ℓ is the jth neuron output at layer ℓ, and $\omega_{j,i}$ is the *synaptic weight* connecting the ith neuron at layer $\ell-1$ with the jth neuron at layer ℓ. Each neuron is activated through a *sigmoid function*. This neural network is trained with the *back-propagation algorithm*, which is based on iteratively using the *gradient descent* to adjust the network weights according to $\omega_{new} = \omega_{old} - \eta\frac{\partial\epsilon}{\partial\omega}$, where ϵ is the *energy function* to be minimized. Adjustment of the network parameters is accomplished backwards from the output layer to the input layer. The architecture of the network, for PD detection, consisted of (1) one hidden layer with seven neurons; (2) one output node corresponding to the two different classes (PD patient, healthy subject); (3) the *logistic function* was used as activation function at each neuron $f(x) = \frac{1}{1+e^{-x}}$; (4) a *learning rate* of 0.1; and (5) 8–20 input nodes, where the number of input nodes was equal to the number of selected features. The number of hidden neurons was selected by gradually adjusting this number from 5 to 20 and testing the corresponding architecture.

The *support vector machine* (SVM) is an optimal classifier, which is geometrically represented by a separating hyper-

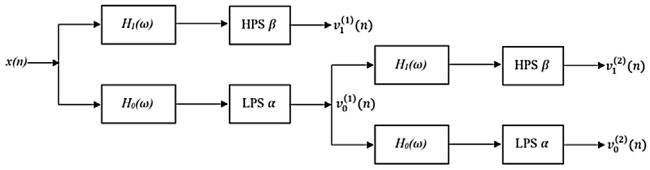

Fig. 3 – Low-pass scaling with $\alpha<1$ (left) and high-pass scaling with $\beta<1$ (right).

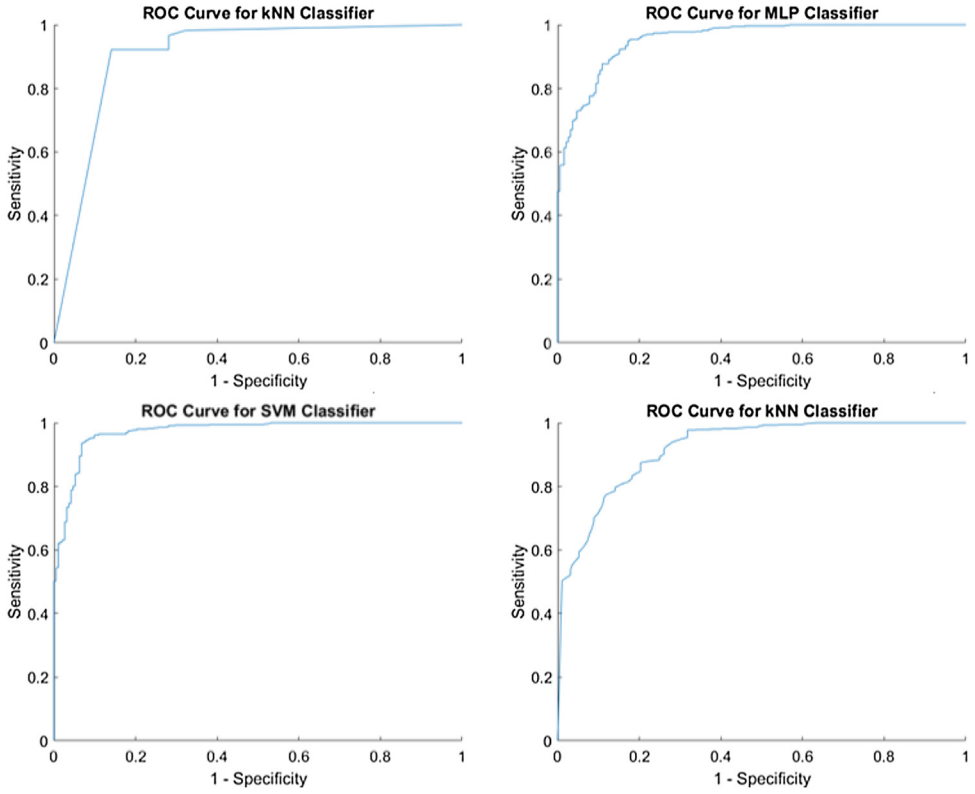

Fig. 4 – Wavelet filter bank. Each stage consists of two channels.

plane which is the furthest away from each class according to $f(\mathbf{x}) = \sum_{i=1}^{n} \omega_i K(\mathbf{x}_i, \mathbf{x}) + b$; where $\mathbf{x}$ is the *feature vector* to be classified, $\mathbf{x}_i$ is the ith *support vector*, ω_i is a coefficient to be estimated, b is the *intercept* of the hyper-plane, and $K(\mathbf{x}_i, \mathbf{x})$ is a *kernel* function. The kernel, used in this work, was a *Radial Basis function*, $K(\mathbf{x}_i, \mathbf{x}) = e^{-\frac{1}{2\sigma^2}||\mathbf{x}-\mathbf{x}_i||^2}$ (σ = 0.1) with (1) one output node corresponding to the two classes of interest, and (2) 8–20 input nodes. The *sequential minimal optimization* (SMO) algorithm is used to solve the dual problem for the derivation of the SVM. The convergence of this algorithm is tested by checking whether the *Karush–Kuhn–Tucker* (KKT) conditions are satisfied to within a tolerance value set to 0.1.

A *decision tree* consists in segmenting the feature space into multiple regions. Region splitting is represented as a tree, where the segmented regions correspond to *leaves*. *Internal nodes* represent hyper-planes where the feature space is split. *Branches* are segments that connect nodes. Recursive binary splitting is used to build a decision tree. Decision trees have high variance, which implies that region splitting could be quite different for different partitions of the training set. *Bootstrap* or *bagging* is a procedure for reducing variance, where multiple decision trees are built using different bootstrapped training sets, and the *tree classifications* are averaged. For a given observation, the assigned class is the most occurring class among multiple tree classifications. In a *random forest* algorithm, a majority of the classifications is not considered at each split in the tree by forcing each split to consider only a subset of the features. The RF implementation consisted of a set of 87–120 decision trees by reaching a tree depth of 150.

2.5. *Performance metrics*

During testing of the method for PD detection, if a PD patient is correctly identified then this case is a *TRUE POSITIVE* (TP); otherwise, this case corresponds to a *FALSE NEGATIVE* (FN). Healthy subjects, correctly diagnosed, correspond to *TRUE NEGATIVES* (TN), while misclassified healthy subjects correspond to *FALSE POSITIVES* (FP).

To evaluate the performance of the proposed method, *accuracy*, *sensitivity*, *specificity*, *precision*, *false alarm rate*, *Matthew correlation coefficient*, F_1 *score* and the *receiver operating curve* (ROC) are used as figures of merit. *Accuracy* is the percentage of diagnosis outcomes that are correct, $Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$. *Sensitivity* or *recall* is the probability that the outcome of diagnosing PD is positive given that the subjects has PD, $Sensitivity = \frac{TP}{TP+FN}$. *Specificity* is defined as the probability that the outcome of PD is negative given that the subject is healthy, $Specificity = \frac{TN}{TN+FP}$. *Precision* is the probability that the outcome of diagnosing PD is true, $Precision = \frac{TP}{TP+FP}$. The *false alarm rate* (FAR) is the probability that the outcome of diagnosing PD is false, $FAR = \frac{FP}{TP+FP}$. The *Matthews Correlation Coefficient* (MCC) is a correlation coefficient between the observed and predicted binary classifications. It returns a value between −1 and +1, $MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$. The F_1 *score* is the harmonic mean of precision and recall, $F_1 = 2\frac{precision \times recall}{precision + recall}$. The *receiver operating characteristic* (ROC) curve is the plot of the Sensitivity against the false positive rate (FPR = 1 − Specificity) in a binary classifier when its threshold is varied.

2.6. Training and testing sets

Feature vectors, from PD or healthy subjects, are stored into two sets: C_1 for patients with PD, C_2 for healthy subjects. Each set, C_1 and C_2, is separated into ten fragments, $C_1 = \{c_{1,1}, c_{1,2}, \ldots, c_{1,10}\}$ and $C_2 = \{c_{2,1}, c_{2,2}, \ldots, c_{2,10}\}$. Then, a fragment $c_{1,i}$ (from C_1) and a corresponding fragment $c_{2,i}$ (from C_2) are randomly combined into c_i. The result of these random mixings is ten fragments or *folds* $\{c_1, c_2, \ldots, c_{10}\}$, where each fold contains instances form PD and healthy subjects. Thus, the set is divided into *ten folds*. One fold is picked for testing of a classifier and the other nine folds are left for training. This process, of choosing one fold for testing and the rest for training, is repeated ten times (ten-fold cross-validation). This process results in ten estimates of a performance metric, $performance_1, \ldots, performance_{10}$. The total estimate is computed by averaging these values.

3. Results

A set of 754 features is fed to the Wrappers selection algorithm. Different signal processing techniques were used to extract these 754 features from the *spectograms* of speech signals during construction of the dataset. The signal processing techniques correspond to different feature groups (baseline features, time-frequency features, MFCCs, WT features, vocal fold features, and TQWT features). Table 1 presents the description of each feature group and the corresponding number of features within each group.

Table 2 presents a description of the selected features, after using Wrappers with each classifier. For each classifier, the size of the selected feature set is specified in the last row, and it is observed that the total number of features ranges from 8 to 20. There were two cases where one baseline feature was selected by Wrappers: when it was run with random forest (*jitter*) and when it was used with the SVM (*DFA*). *Model selection*, to determine the optimal architecture parameters for each classifier, was executed along with feature subset selection.

The computational time, to select the feature set and classifier parameters, depended on the number of model parameters, with 72 h for feature selection based on a MLP classifier, 27 h for feature selection based on RF, 23 h for SVM, and 10 h for kNN. These sets of experiments were run in a processor Intel core i5-8250U at 1.60 GHz with 6 MB of cache memory, 4 Cores and 8 threads.

Table 3 shows the PD detection performance in four different classifiers, (kNN, MLP, SVM and RF) when they are tested with a feature set selected by running the Wrappers algorithm with a kNN classifier. The best results are highlighted in boldface. Tables 4–6 show the PD detection performance of the four classifiers, when they are tested with a feature set selected by running the Wrappers algorithm with MLP, SVM and RF, respectively. These results were obtained by using 10-fold cross-validation.

The results, shown in Tables 3–6, were obtained by using a MLP implementation based on a learning rate of 0.1 and 2500 epochs, a kNN implementation based on the three nearest neighbors ($k = 3$), a RF implementation based on 87–120 decision trees with a tree depth of 150, and a SVM-RBF implementation based on tolerance values of 0.05 and 0.1, and cost values of 15 and 20.

According to Tables 3–6, the best performance metric values for vocal-based PD detection are *Accuracy* = 0.947, *Sensitivity* = 0.984, *Specificity* = 0.9268, *Precision* = 0.9722, *False Alarm Rate* = 0.0277, *MCC* = 0.8686, F_1 *score* = 0.9633. Plots, of the average ROC curves for four classifiers tested with a feature set selected by Wrappers with a SVM classifier, are shown in Fig. 5.

Although a direct comparison of our method with current approaches is difficult, due to the use of different datasets, some results on PD detection are shown in Table 7. The purpose of Table 7 is to show different methodologies, number of selected features and performance results, proposed by the scientific community, working on PD detection.

4. Discussion

In this work, different vocal feature subsets were merged to select an optimal subset of features, with attributes from four different subsets, including the most commonly used set in PD

Table 1 – Features extracted from spectograms of speech signals.

Feature set	Description	Total features
Baseline features	Jitter and shimmer variants. Mean, median, standard deviation, minimum and maximum values of the vocal fold vibration frequencies. Harmonicity parameters. Recurrence period density entropy, detrended fluctuation analysis, and pitch period entropy.	32
Time-frequency features	Mean, minimum and maximum speech signal power in dB. First four formant frequencies in the vocal tract. First four bandwidths of formant frequencies.	11
MFCCs	Mean and standard deviation of the original 13 MFCCs and their first-second derivatives. Log-energy of the signal.	84
Wavelet transform	Discrete wavelet transform is extracted at ten levels. The Shannon's and the log energy entropy are computed. The Teager–Kaiser energy of the approximation and detailed coefficients is computed.	182
Vocal fold features	Periodicity of glottis movements. Glottal to Noise Excitation. Amount of noise due to the pathological vocal fold vibration. Empirical Mode Decomposition.	22
TQWT	The TQWT algorithm relies on Q (*Q-factor* from 1 to 10), r (*redundancy* chosen as 3, 4, 5) and J (*number of levels* between 5 and 50) parameters. After decompositions with the TQWT, energy/entropy values of each level were computed.	423

Table 2 – Features obtained by using Wrappers with each classifier (dec. stands for *decomposition* and std. stands for *standard deviation*).

Feature subset	Features			
	kNN	MLP	SVM	RF
Baseline features	0	0	DFA	Jitter
MFCCs	Std. double-delta energy	Mean 3rd delta, std. double-delta energy	Mean 6th coef., std. 7th delta, std. 9th delta, std. double-delta energy	Std. 1st double-delta, std. 9th double-delta
WT features	0	3rd coef.	Entropy 1st coef., entropy 4th coef., entropy-log 2nd coef.	Entropy 3rd coef., entropy-log 8th coef., 1st coef., mean 5th coef.
TQWT features	Energy dec. 35, mean dec. 12, std. dec. 28, std. dec. 16, min. dec. 17, skewness dec. 24, kurtosis dec. 8	Entropy dec. 12, mean dec. 12, std. dec. 6, std. dec. 13, std. dec. 17, min. dec. 13, min. dec. 17, min. dec. 35, skewness dec. 17	Energy dec. 6, entropy dec. 35, entropy-log dec. 27, entropy-log dec. 31, entropy-log dec. 32, entropy-log dec. 33, mean dec. 18, mean dec. 25, std. dec. 12, kurtosis dec. 26, kurtosis dec. 31, kurtosis dec. 36	Energy dec. 11, energy dec. 14, entropy dec. 32, entropy-log dec. 1, entropy-log dec. 28, entropy-log dec. 32, entropy-log dec. 33, entropy-log dec. 34, entropy-log dec. 35, mean dec. 12, std. dec. 5
Number of features	8	12	20	18

Table 3 – PD-detection performance metrics for four different classifiers by using a feature set selected by Wrappers with a kNN classifier.

	Accuracy	Sensitivity	Specificity	Precision	FAR	MCC	F_1 score
kNN	**0.8558**	0.9379	**0.6146**	**0.8773**	**0.1227**	**0.5986**	0.9066
MLP	**0.8558**	0.9521	0.5729	0.8675	0.1325	0.5933	**0.9079**
SVM-RBF	0.836	**0.9699**	0.4427	0.8364	0.1636	0.5256	0.8982
RF	0.8479	0.9309	0.6042	0.8735	0.1265	0.5768	0.9013

Table 4 – PD detection performance for four different classifiers by using a feature set selected by Wrappers with a MLP classifier.

	Accuracy	Sensitivity	Specificity	Precision	FAR	MCC	F_1 score
kNN	0.8201	0.9167	0.5365	0.8531	0.1469	0.4946	0.8838
MLP	**0.8664**	0.9362	**0.6615**	**0.8904**	**0.1096**	**0.6326**	**0.9127**
SVM-RBF	0.8307	**0.9557**	0.4635	0.8396	0.1604	0.5099	0.8939
RF	0.8479	0.9433	0.5677	0.865	0.135	0.571	0.9025

Table 5 – PD detection performance for four different classifiers by using a feature set selected by Wrappers with a SVM classifier.

	Accuracy	Sensitivity	Specificity	Precision	FAR	MCC	F_1 score
kNN	0.9034	0.9663	0.7188	0.9098	0.0901	0.735	0.9064
MLP	0.9206	0.9592	0.8073	0.935	0.064	0.786	0.9284
SVM-RBF	**0.947**	0.9645	**0.9268**	**0.9722**	**0.0277**	**0.8686**	**0.9633**
RF	0.9021	**0.977**	0.6823	0.9003	0.0996	0.735	0.9011

Table 6 – PD detection performance for four different classifiers by using a feature set selected by Wrappers with a RF classifier.

	Accuracy	Sensitivity	Specificity	Precision	FAR	MCC	F_1 score
kNN	0.9101	0.9681	0.7396	0.9161	0.0839	0.7541	0.9414
MLP	0.8995	0.9468	**0.7604**	**0.9207**	**0.0793**	0.7284	0.9336
SVM-RBF	0.8743	0.977	0.5729	0.8705	0.1295	0.6485	0.9206
RF	**0.922**	**0.984**	0.7396	0.9174	0.0826	**0.7878**	**0.9495**

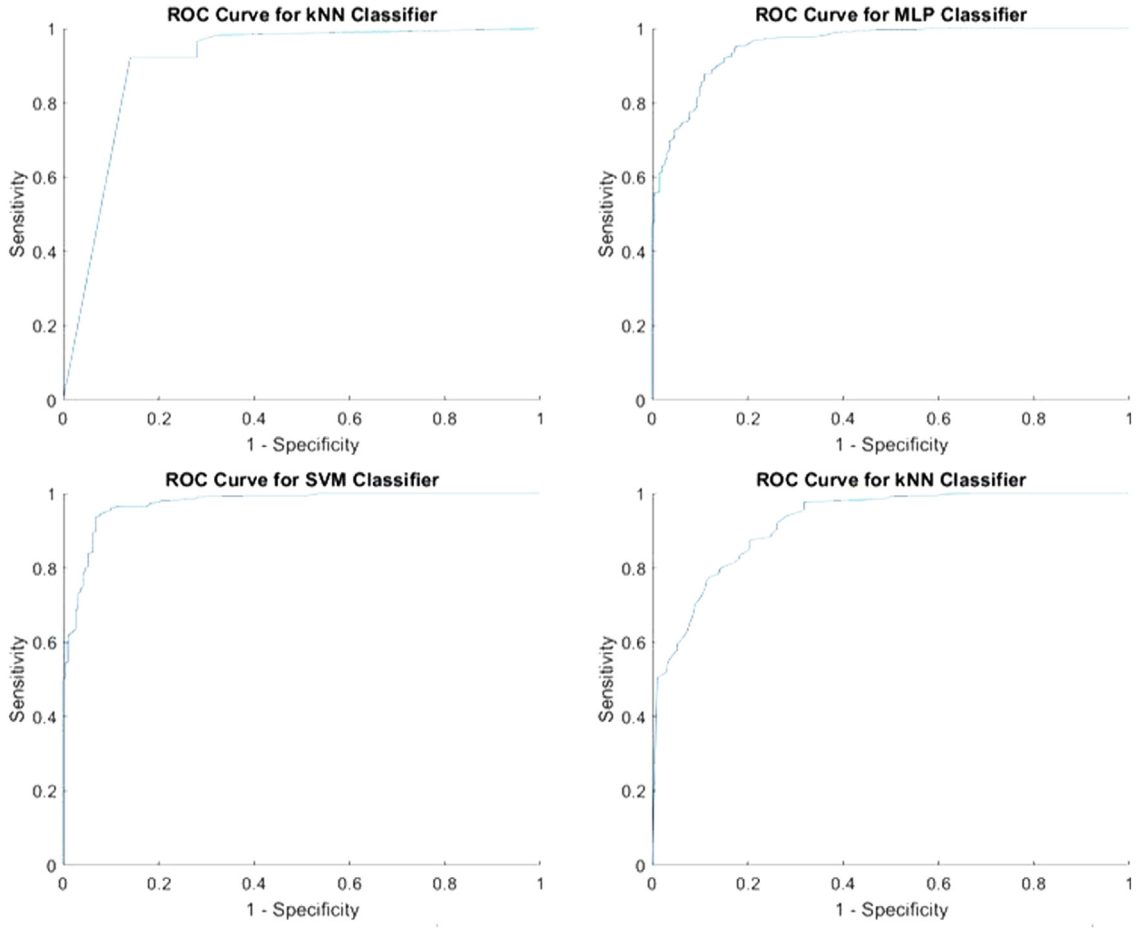

Fig. 5 – ROC curve for four classifiers using 10-fold CV.

detection, referred to as *baseline feature set*, which provides only one feature in those experiments where Wrappers is run with the SVM and RF classifiers. According to Table 1, the TQWT feature subset is the one with the highest contribution since these features carry significant discriminative information in PD detection, a conclusion that is similar to that reported by Sakar et al. [13]. MFCCs provide the second-highest contribution. It has been previously shown that MFCCs and TQWT coefficients contain complementary information that provide higher classification accuracy when they are used together [13]. Wavelet transform (WT) features provide the third highest-contribution. In the first work to use the same dataset [13], six feature subsets were selected after using *minimum-redundancy maximum-relevance*, where *TQWT* is the most relevant subset with 30 features, *MFCC* is the second top subset with 10 features; followed by the *Baseline* set, *Vocal fold* set, *Bandwidth* set and *WT* set, with 4, 3, 2 and 1 features, respectively.

The extraction of WT, MFCC and TQWT features is based on the use of filter banks, where the bandwidth of each filter is proportional to the central frequency. Filters, at lower frequency bands, are characterized by lower bandwidths; while filters, at higher frequency bands, are characterized by higher bandwidths. This behavior is similar to the way the human auditory system perceives voice where small changes in low frequency components are finely perceived. In contrast, the auditory system cannot discern similar differences between the two closely spaced frequency components at higher frequencies. A larger frequency separation between components is required at higher frequencies.

An independent analysis of each table, from 3 to 6, shows that the performance of one particular classifier is better than that of the others when the feature set is obtained by running the Wrappers algorithm with the same particular classifier. Furthermore, by comparing all the results in Tables 3–6, it is observed that the best performance for the kNN, MLP and SVM is obtained when these classifiers are tested with a feature set obtained by running Wrappers with the SVM classifier. The RF classifier reaches the highest PD detection performance when it is tested with a feature set selected by running Wrappers with RF. According to Table 7, the most applied classifier to the problem of vocal-based PD detection is SVM.

All the best metric values, but one, correspond to the PD detection based on the SVM classifier when it is tested with a set selected by running Wrappers with a SVM. Thus, SVM tend to perform better than the other classifiers for the application of vocal-based PD detection.

According to Table 7, other works on vocal-based PD detection have reported higher accuracy than the best accuracy obtained in this study; however, all of these studies

Table 7 – Some methods for Parkinson detection.

Author, Year	Detection algorithm	Number of features	Results
Sakar et al. (2019) [13]	Naive Bayes, Logistic Regression, kNN, MLP, RF, SVM-Linear, SVM-RBF, Baseline Features, Wavelet features, MFCC, TQWT	50 features	Accuracy = 0.86, F_1 score = 0.84, MCC = 0.59
Peker (2016) [5]	SVM, baseline features	22 features from 195 sound measurements from 8 healthy people and 23 PD patients	Accuracy = 0.99, Sensitivity = 0.96, Specificity = 1, F_1 score = 0.98
Guruler (2017) [9]	ANN and Complex-Valued ANN (CVANN), K-means clustering-based feature weighting, baseline features	22 features from 195 sound measurements from 8 healthy subjects and 23 with PD	Accuracy = 0.99, Sensitivity = 1, Specificity = 0.99, F_1 score = 0.99
Sakar et al. (2017) [4]	kNN, SVM, Extreme Learning Machine (ELM), baseline features	16 features from 42 with PD	Accuracy = 0.96, MCC = 0.77
Braga et al. (2019) [2]	ANN, SVM, RF, baseline features	19 features from 22 speakers with PD and 30 healthy speakers	Accuracy = 0.99 for RF classifier
Sakar et al. (2013) [3]	kNN, SVM-linear, SVM-RBF, linear and time-frequency based features	26 features belong to 20 patients with PD and 20 healthy individuals	Accuracy = 0.85, Sensitivity = 0.85, Specificity = 0.9
Peker et al. (2015) [17]	mRMR feature subset selection, CVANN, baseline features	22 features from 195 sound measurements from 23 PD patients and 8 healthy people	Accuracy = 1, Sensitivity = 1, Specificity = 1, F_1 score = 0.99
Upadhya et al. (2018) [29]	MLP, baseline features, MFCCs	40 features	Accuracy = 0.98, Sensitivity = 0.99, Specificity = 0.99
Tsanas et al. (2011) [10]	SVM, RF, baseline features	132 features from 10 healthy controls and 33 patients with PD	Accuracy = 0.977
Proposed approach	kNN, MLP, SVM, RF, baseline features, MFCC, WT features, TQWT features	8 to 20 features from 64 healthy individuals and 188 patients with PD	Accuracy = 0.947, Sensitivity = 0.984, Specificity = 0.9268, Precision = 0.9722, False Alarm Rate = 0.0277, MCC = 0.8686, F_1 score = 0.9633

used a smaller dataset and these studies use leave-one-out cross validation technique, which results in biased results.

The reported datasets consist of collections of voice instances, which were extracted from 20 PD patients (20 healthy subjects) [3], 22 PD speakers (30 healthy speakers) [2,30], 23 subjects with PD (8 subjects are healthy) [5,9,31], 28 PD patients [2,32], 33 PD patients (10 healthy subjects) in the *National Center for Voice and Speech* dataset [33], 42 PD patients (8 healthy subjects) [4,14,34], 45 PD patients (45 subjects are healthy) [29]. It is considered that the number of subjects in these datasets is reduced with the disadvantage that the performance of these methods may degrade when tested on datasets with a higher number of subjects [13]. For the case of our study, to the best of our knowledge, we are using the largest public dataset, which contains data collected from 188 PD patients and a control group of 64 healthy individuals.

The work by Sakar et al. [13] introduces the current dataset. That work is the first to use the TQWT coefficients as features for vocal-based PD detection. According to Table 6, that work reported an *Accuracy* = 0.86, F_1 *score* = 0.84, and *MCC* = 0.59. The method, described in this paper, is characterized by *Accuracy* = 0.947, F_1 *score* = 0.9633, and *MCC* = 0.8686. The work by Sakar et al. (2019) uses a feature set of 50 attributes after running the *mRMR* feature subset selection algorithm whereas the proposed method is based on a set, where the maximum number of features is 20 and the minimum is 8.

5. Conclusions

The problem of vocal-based PD detection at early stages, based on the largest available dataset and feature subset selection, is studied. It is shown that the application of Wrappers subset selection is appropriate because of the low dimensionality of the selected feature set and the fact that the proposed method increases PD detection performance. Evidence of the effectiveness of the proposed method is obtained through assessment of different performance metrics. The accuracy of the proposed method, which is 0.94, is higher than that from the first work to use the same dataset, which is 0.86. The improvement over the previous work is not only in terms of *accuracy*, but also in terms of *complexity* since this performance was obtained by feeding 8–20 features to classifiers, while the latest work is based on 50 features.

It is confirmed that the TQWT features are the group with the highest contribution when it comes to discriminate between patients with PD and healthy subjects. It is found that the most effective classifier for vocal-based PD detection is the SVM-RBF. Furthermore, it is found that the most significant features are extracted by banks of filters, where the bandwidths are proportional to the center frequency. The filtering process is similar to the way the human auditory system analyses voice signals.

Conflict of interest

The authors declare that they do not have conflicts of interest with other people or organizations that could inappropriately influence their work.

REFERENCES

[1] Michael J. Fox Foundation for Parkinson research, Parkinson's disease causes; 2018, Retrieved from: https://www.michaeljfox.org/understanding-parkinsons/living-with-pd.html.

[2] Braga D, Madureira AM, Cuelho L, Ajith R. Automatic detection of Parkinson's disease based on acoustic analysis of speech. Eng Appl Artif Intell 2019;77:148–58. http://dx.doi.org/10.1016/j.engappai.2018.09.018

[3] Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, et al. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inform 2013;19(4):828–34. http://dx.doi.org/10.1109/JBHI.2013.2245674

[4] Sakar BE, Serbes G, Sakar CO. Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease. PLOS ONE 2017;1–18. http://dx.doi.org/10.1371/journal.pone.0182428

[5] Peker M. A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM. J Med Syst 2016;40 (116):1–16. http://dx.doi.org/10.1007/s10916-016-0477-6

[6] Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. A decision support system to improve medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM. IEEE Trans Biomed Eng 2012;59(5):1264–71. http://dx.doi.org/10.1109/TBME.2012.2183367

[7] Zhang HH, Yang L, Liu Y, Wang P, Yin J, Li Y, et al. Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples. Biomed Eng Online 2016;15(122). http://dx.doi.org/10.1186/s12938-016-0242-6

[8] Berus L, Klancnik S, Brezocnik M, Ficko M. Classifying Parkinson's disease based on acoustic measures using artificial neural networks. Sensors 2019;19(1). http://dx.doi.org/10.3390/s19010016

[9] Guruler H. A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. Neural Comput Appl 2017;28:1657–66. http://dx.doi.org/10.1007/s00521-015-2142-2

[10] Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J R Soc Interface 2011;8(59):842–55. http://dx.doi.org/10.1098/rsif.2010.0456

[11] Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomed Eng Online 2007;6(23). http://dx.doi.org/10.1186/1475-925X-6-23

[12] Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng 2009;56 (4):1010–22. http://dx.doi.org/10.1109/TBME.2008.2005954

[13] Sakar CO, Serbes G, Gunduz A, Tunc HC, Nizam H, Sakar BE, et al. A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable q-factor wavelet transform. Appl Soft Comput 2019;74:255–63. http://dx.doi.org/10.1016/j.asoc.2018.10.022

[14] Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Trans Biomed Eng 2010;57 (4):884–93. http://dx.doi.org/10.1109/TBME.2009.2036000

[15] Selesnick IW. Wavelet transform with tunable q-factor. IEEE Trans Signal 2011;59(8):3560–75. http://dx.doi.org/10.1109/TSP.2011.2143711

[16] Selesnick IW. Resonance-based signal decomposition: a new sparsity-enabled signal analysis method. Signal Process 2011;91(12):2793–809. http://dx.doi.org/10.1016/j.sigpro.2010.10.018

[17] Peker M, Sen B, Delen D. Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and MRMR feature selection algorithm. J Healthc Eng 2015;6 (3):281–302. http://dx.doi.org/10.1260/2040-2295.6.3.281

[18] Kohavi R, John GH. Wrappers for feature subset selection. Artif Intell 1997;97:273–324.

[19] Braak H, Ghebremedhin E, Rüb U, Bratzke H, Del Tredici K. Stages in the development of Parkinson's disease-related pathology. Cell Tissue Res 2004;(318):121–34. http://dx.doi.org/10.1007/s00441-004-0956-9

[20] Banga YI, Minc K, Sohnd YH, Choa SR. Acoustic characteristics of vowel sounds in patients with parkinson disease. Neuro Rehabil 2013;32:649–54. http://dx.doi.org/10.3233/NRE-130887

[21] Boersma P. Praat: doing phonetics by computer, ear hear; 2011, http://www.fon.hum.uva.nl/praat/.

[22] Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;8(27):1226–38. http://dx.doi.org/10.1109/TPAMI.2005.159

[23] Jafari GR, Pedram P, Ghafoori Tabrizi K. Detrended fluctuation analysis of Bach's inventions and sinfonias pitches. AIP Conf Proc 2007;889(1). http://dx.doi.org/10.1063/1.2713472

[24] Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL. Mosaic organization of DNA nucleotides. Phys Rev E 1994;49(2).

[25] Muda L, Begam M, Elamvazuthi I. Voice recognition algorithms using Mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. J Comput 2010;2(3):674–93.

[26] Sigurdsson S, Petersen KB, Lehn-Schiøler T. Mel frequency cepstral coefficients: an evaluation of robustness of MP3 encoded music. ISMIR 2006.

[27] Picone J. Signal modeling techniques in speech recognition. Proc IEEE 1993;81(9):1215–47.

[28] Mallat S. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 1989;11(7):674–93.

[29] Upadhya SS, Sheeran AN. Discriminating Parkinson and healthy people using phonation and cepstral features of speech. Proc Comput Sci 2018;143:197–202. http://dx.doi.org/10.1016/j.procs.2018.10.376

[30] Proença J, Perdigão F, Veira A, Candeias S, Lemos J, Januário C. Characterizing Parkinson's disease speech by acoustic

and phonetic features. Lect Notes Comput Sci 2014;8775:24–35.

[31] Bache K, Lichman M. UCI machine learning repository; 2013, Available at http://archive.ics.uci.edu/ml.

[32] UCI. UCI Machine Learning Repository: Parkinson speech dataset with multiple types of sound recordings data set; 2014, Available at: https://archive.ics.uci.edu/ml/machine-learning-databases/00470/.

[33] 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. http://dx.doi.org/10.1109/TBME.2012.2183367

[34] Grover S, Bhartia S, Akshama, Yadav A, Seeja KR. Predicting severity of Parkinson's disease using deep learning. Proc Comput Sci 2018;132:1788–94. http://dx.doi.org/10.1016/j.procs.2018.05.154