http://doi.org/10.35784/iapgos.3685 *received: 29.05.2023 | revised: 19.07.2023 | accepted: 25.09.2023 | available online: 30.09.2023*

CLASSIFICATION OF PARKINSON'S DISEASE AND OTHER NEUROLOGICAL DISORDERS USING VOICE FEATURES EXTRACTION AND REDUCTION TECHNIQUES

Oumaima Majdoubi, Achraf Benba, Ahmed Hammouch
Mohammed V University in Rabat, National School of Arts and Crafts, Electronic Systems Sensors and Nanobiotechnology, Rabat, Morocco

Abstract. *This study aimed to differentiate individuals with Parkinson's disease (PD) from those with other neurological disorders (ND) by analyzing voice samples, considering the association between voice disorders and PD. Voice samples were collected from 76 participants using different recording devices and conditions, with participants instructed to sustain the vowel /a/ comfortably. PRAAT software was employed to extract features including autocorrelation (AC), cross-correlation (CC), and Mel frequency cepstral coefficients (MFCC) from the voice samples. Principal component analysis (PCA) was utilized to reduce the dimensionality of the features. Classification Tree (CT), Logistic Regression, Naive Bayes (NB), Support Vector Machines (SVM), and Ensemble methods were employed as supervised machine learning techniques for classification. Each method provided distinct strengths and characteristics, facilitating a comprehensive evaluation of their effectiveness in distinguishing PD patients from individuals with other neurological disorders. The Naive Bayes kernel, using seven PCA-derived components, achieved the highest accuracy rate of 86.84% among the tested classification methods. It is worth noting that classifier performance may vary based on the dataset and specific characteristics of the voice samples. In conclusion, this study demonstrated the potential of voice analysis as a diagnostic tool for distinguishing PD patients from individuals with other neurological disorders. By employing a variety of voice analysis techniques and utilizing different machine learning algorithms, including Classification Tree, Logistic Regression, Naive Bayes, Support Vector Machines, and Ensemble methods, a notable accuracy rate was attained. However, further research and validation using larger datasets are required to consolidate and generalize these findings for future clinical applications.*

Keywords: voice analysis, Parkinson's disease, MFCC, PCA, naive Bayes kernel, machine learning

KLASYFIKACJA CHOROBY PARKINSONA I INNYCH ZABURZEŃ NEUROLOGICZNYCH Z WYKORZYSTANIEM EKSTRAKCJI CECH GŁOSOWYCH I TECHNIK REDUKCJI

Streszczenie. *Przedstawione badanie miało na celu różnicowanie osób z chorobą Parkinsona (PD) od osób z innymi zaburzeniami neurologicznymi poprzez analizę próbek głosowych, biorąc pod uwagę związek między zaburzeniami głosu a PD. Próbki głosowe zostały zebrane od 76 uczestników przy użyciu różnych urządzeń i warunków nagrywania, a uczestnicy byli instruowani, aby wydłużyć samogłoskę /a/ w wygodnym tempie. Oprogramowanie PRAAT zostało zastosowane do ekstrakcji cech, takich jak autokorelacja (AC), krzyżowa korelacja (CC) i współczynniki cepstralne Mel (MFCC) z próbek głosowych. Analiza składowych głównych (PCA) została wykorzystana w celu zmniejszenia wymiarowości cech. Jako techniki nadzorowanego uczenia maszynowego wykorzystano drzewa decyzyjne (CT), regresję logistyczną, naiwny klasyfikator Bayesa (NB), maszyny wektorów nośnych (SVM) oraz metody zespołowe. Każda z tych metod posiadała swoje unikalne mocne strony i charakterystyki, umożliwiając kompleksową ocenę ich skuteczności w rozróżnianiu pacjentów z PD od osób z innymi zaburzeniami neurologicznymi. Naiwny klasyfikator Bayesa, wykorzystujący siedem składowych PCA, osiągnął najwyższy wskaźnik dokładności na poziomie 86,84% wśród przetestowanych metod klasyfikacji. Należy jednak zauważyć, że wydajność klasyfikatora może się różnić w zależności od zbioru danych i konkretnych cech próbek głosowych. Podsumowując, to badanie wykazało potencjał analizy głosu jako narzędzia diagnostycznego do rozróżniania pacjentów z PD od osób z innymi zaburzeniami neurologicznymi. Poprzez zastosowanie różnych technik analizy głosu i wykorzystanie różnych algorytmów uczenia maszynowego, takich jak drzewa decyzyjne, regresja logistyczna, naiwny klasyfikator Bayesa, maszyny wektorów nośnych i metody zespołowe, osiągnięto znaczący poziom dokładności. Niemniej jednak, konieczne są dalsze badania i walidacja na większych zbiorach danych w celu skonsolidowania i uogólnienia tych wyników dla przyszłych zastosowań klinicznych.*

Słowa kluczowe: analiza głosu, choroba Parkinsona, MFCC, PCA, naiwne jądro bayesowskie, uczenie maszynowe

Introduction

Parkinson's disease, which was initially described by Dr. James Parkinson in 1817, is a progressive nervous system condition that deteriorates over time and affects physical movements, including speech [23]. It is the second most prevalent neurological disorder, surpassed only by Alzheimer's, Multiple System Atrophy (MSA), brain tumors, epilepsy, and other neurodegenerative conditions. The disease is caused by the degradation of pigment cells in the basal ganglia, resulting in a deficiency of dopamine and disruption of neurotransmission in the Substantia Nigra region of the midbrain, which is responsible for regulating motor function. Motor problems result from dopamine insufficiency. Symptoms include bradykinesia, postural instability, tremors, and hypokinetic movement abnormalities. Although this condition is straightforward to diagnose in its advanced stages, effective therapy is difficult [22]. There is presently no curative medicinal therapy for Parkinson's disease.

The study of voice abnormalities associated with neurological conditions, which can be caused by various factors that affect muscle tone [29], has significantly advanced. These conditions are categorized as either hypotonia, which refers to low muscle tone, or hypertonia, which refers to high muscle tone. Hypotonia is characterized by reduced loudness, changes in fundamental frequency, and voice instability, while hypertonia is characterized by vocal pauses, voice instability [3], and changes in voice quality [27]. Neurological conditions such as vocal tremors, spasmodic dysphonia, and vocal cord paralysis can all impair the voice. Moreover, Parkinson's disease can impact speech and voice, leading to diminished pitch and loudness fluctuations, decreased overall loudness, and a breathy quality in the voice due to insufficient closure of the vocal cords. Subjective approaches are routinely used by clinicians and vocal pathologists to evaluate speech impairments in Parkinson's disease patients, which can have a substantial impact on communication and quality of life. These approaches utilize acoustic characteristics including fundamental frequency, sound intensity level, tremor, shimmer, ratio of low-frequency to high-frequency components, cephalic peak prominence, and harmonics-to-noise ratio within the signal for assessment [1, 20, 31]. In contrast, recent research has shifted towards utilizing acoustic parameters derived from time-based data, along with spectrum and cepstral measurements, for more objective evaluations [2]. These evaluations provide a more precise understanding of the severity and characteristics of speech disorders in Parkinson's disease (PD) patients and are crucial for effective management. PD commonly results in voice weakening in about 90% of patients [21], typically occurring in those over 60 years old, although it can also affect younger individuals, albeit rarely. Additionally, gender appears to play a role in PD prevalence, with men being more affected than women [10]. Recent research has focused on developing objective methods to diagnose vocal problems by measuring voice quality in the temporal, spectral, and cepstral domains, including parameters such as fundamental frequency (F0), absolute sound pressure level, jitter, shimmer, and harmonicity [18, 19, 25].

In [12], the authors proposed a method for diagnosing PD using speech signals. They collected a dataset of speech recordings from individuals with PD and healthy controls. The method involved extracting resonance and time-frequency-based features from the speech signals, but the paper lacked specific details on the algorithms and parameters used. The extracted features were fused using an undisclosed technique, and machine learning classifiers were employed for PD diagnosis. However, the paper did not specify the classifiers used or provide information on the training and evaluation procedures. The paper also lacked crucial details regarding the dataset, feature extraction, fusion technique, classifiers, and performance evaluation metrics. A comprehensive discussion of the results and their interpretation was also missing. Addressing these limitations would improve the study's scientific rigor and impact.

The authors in [24] presented an approach to diagnose PD by utilizing cepstral features extracted through MFCC and dimensionality reduction techniques. However, the paper did not provide sufficient information regarding the dataset, feature extraction parameters, dimensionality reduction methods, SVM classifier parameters, and performance evaluation metrics. Furthermore, the study lacked in-depth discussions and interpretations of the results, which diminished its overall strength. To improve the research's scientific rigor and impact, it is crucial to address these limitations and provide comprehensive details and analysis.

The primary focus of the study described in [26] is the investigation of feature extraction and classification techniques specifically designed for dysphonic speech disorder in individuals with PD. The main objective is to identify effective methods for extracting pertinent features from dysphonic speech signals and to compare different classification techniques to achieve accurate diagnosis of PD-related dysphonia. The paper thoroughly examines various feature extraction techniques and provides detailed information on the algorithms and parameters employed. Furthermore, it assesses different classification methods using performance metrics. The findings of this study contribute to enhancing our understanding of dysphonic speech in PD and offer valuable insights into the diagnosis of PD-related dysphonia.

While most studies focus on distinguishing between patients with Parkinson's disease and healthy individuals, our research stands out by focusing on classifying Parkinson's disease and other neurological disorders. The objective of our study is to provide an objective method for diagnosing vocal problems in Parkinson's disease patients and differentiating them from patients with other neurodegenerative diseases. By collecting voice recordings from patients with Parkinson's disease, Multiple System Atrophy, and other neurological disorders, we built a database of 76 voice samples. Through the utilization of cepstral domains, we extracted acoustic features from each voice sample, and principal component analysis (PCA) was employed to select the most relevant features. We then employed various supervised machine-learning techniques for classification. The combination of the Naive Bayes kernel with linear classification and 7 components of PCA achieved a maximum classification accuracy of 86.84%. This machine learning method, along with artificial intelligence techniques, can aid in the detection of Parkinson's disease, Multiple System Atrophy, and other neurological diseases. Once the algorithm is trained to recognize these patterns, it can effectively classify patients based on their health status, distinguishing Parkinson's disease patients from those with similar symptoms but other neurodegenerative disorders. Our study has the potential to enhance the accuracy of diagnosis, leading to improved management and treatment of Parkinson's disease.

Unlike [12, 18, 19, 25], the significance of evaluating acoustic characteristics for Parkinson's disease diagnosis is emphasized in this study, regardless of whether dimension reduction is employed. Certain features have exhibited high accuracy in distinguishing Parkinson's disease from other disorders. Additionally, another study utilizing dimension reduction has also shown promising outcomes.

In contrast to previous studies [24, 26], the comparison of various machine learning classifiers such as Classification Tree (CT), Logistic Regression, Naive Bayes (NB), Support Vector Machines (SVM), and Ensemble methods reveals their effectiveness in accurately detecting and distinguishing between patient groups, thereby assisting in clinical diagnosis.

To the best of our knowledge, this study represents the only existing research focused on classifying patients with Parkinson's disease from other neurological diseases with similar symptoms.

Furthermore, our study goes beyond previous research by considering a larger and more diverse dataset, allowing for a more comprehensive analysis of the classification between patients with Parkinson's disease and those with other neurological disorders.

The paper is organized as follows: Section 1 presents the methodology and database used in this study, Section 2 discusses the evaluation metrics, Section 3 presents the obtained results and discussion, and Section 4 provides the conclusion.

1. Methodology

In this research, we utilized the cepstral technique to extract cepstral coefficients from acoustic recordings. We then employed PRAAT software to analyze and extract features, specifically Mel-frequency cepstral coefficients (MFCC), from these samples. This section provides a comprehensive overview of the complete procedure involved in these three approaches, along with their categorization.

1.1. Dataset

All participants in the study provided self-reported diagnoses of Parkinson's disease, MSA, or other ND. Voice recordings were collected using two methods: some participants used their own devices to record their voices, while others were recorded by physicians using a smartphone microphone. The study utilized a database obtained from a previous study [4], which consisted of 76 voice samples. Out of these, 56 samples were collected from Parkinson's disease patients, who were divided into three groups based on the recording method: the first group used smartphones (10 females and 22 males), the second group used tablets (5 females and 3 males), and the third group used a computer (6 females and 10 males). Additionally, voice recordings were obtained from 20 patients with various other neurological disorders, including Multiple System Atrophy 9 patients (2 females and 7 males), Functional Neurological Disorders 5 patients (3 females and 2 males), Essential Tremor (1 female), Dystonia (2 males and 1 female), Cervical Dystonia (1 female), and Somatization (1 female). All voice samples were saved in WAV format and recorded in mono-channel mode.

1.2. Feature extraction

Some individuals with vocal cord pathology may struggle to maintain stable phonation when pronouncing a sustained vowel /a/ [5]. Voice recordings and pre-processing alone may not be sufficient to accurately assess voice disorders. Therefore, we opted to utilize the acoustic software PRAAT to extract a comprehensive set of information on speech characteristics. We employed two methods: autocorrelation (all parameters) to pitch (ac), and cross-correlation method to pitch (cc). A total of 73 parameters were extracted for speech analysis, including 10 jitter parameters (dc and ac) [jitter (local), jitter (local, absolute), jitter (rap), jitter (ppq5), jitter (ddp)], 12 speckle parameters (dc and ac) [(speckle (local), speckle (local, dB), speckle (apq3), speckle (apq5), speckle (apq11), speckle (dda)], 6 harmonicity parameters (dc and ac) (mean autocorrelation, mean noise/harmonic ratio, mean harmonic/noise ratio), 10 pitch parameters (cc and ac) (median pitch, mean pitch, standard deviation, minimum pitch, maximum pitch), 8 pulse parameters (cc and ac) (number of pulses, number of periods, mean period, standard deviation of period), 6 vocalization parameters (fraction of locally unvoiced

pitch frames, number of voicing breaks, degree of voicing breaks), frequency of formants 1, 2, and 3, maximum, minimum, and mean intensity, as well as 15 first coefficient of MFCC.

Jitter and Shimmer are significant metrics used in the evaluation of vocal cord pathology in patients who are affected. Jitter measures the variation in the fundamental frequency (F0) of the voice from cycle to cycle [16, 17], while Shimmer measures the variation in voice cycle width [32]. The algorithm used to calculate Jitter and Shimmer involves computing the mean of the disparities in duration or amplitude or amplitude of consecutive periods. These measurements are useful in aiding the diagnosis and monitoring of vocal cord pathology, as well as evaluating the effectiveness of treatments.

PCA (Principal Component Analysis) is a popular statistical approach for reducing the dimensionality of voice data and extracting the most important information. After applying PCA to the speech data, we obtained 13 principal components, ranging from the first principal component, which accounts for the highest variance, to the 13th principal component, which captures the least variance. The basic idea behind PCA is to describe the variance in a multidimensional dataset using a collection of uncorrelated variables that are linear combinations of the original variables. These new variables, dubbed "principal components", are arranged in decreasing order of significance, with the first representing the most variance in the original data. Subsequent components are chosen to have the least amount of variance while staying uncorrelated with the prior components.

1.2.1. Jitter measurement

Jitter is a parameter that measures micro-perturbation of the fundamental period and has several parameters [28]:

- Jitter (relative or local) refers to the timing variations or fluctuations of a signal relative to a reference signal or a clock, expressed as the difference in timing between the expected timing and the actual timing of signal transitions with respect to a reference signal [11]. Is calculated as follows:

$$\text{Jitter (relative)} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} T_i}$$

Where Ti represents the period lengths of the extracted fundamental frequency $F0$ s and N is denotes the total number of extracted $F0$ periods.

Table 1. Time-frequency-based features given by Praat acoustic analysis software

Groups	Features
Pitch parameters	Median pitch (Hz)
	Mean pitch (Hz)
	Standard deviation (Hz)
	Minimum pitch (Hz)
	Maximum pitch (Hz)
Pulses parameters	Number of pulses
	Number of periods
	Mean period (s)
	Standard deviation of period (s)
Voicing parameters	Fraction of locally unvoiced frames (%)
	Number of voice breaks
	Degree of voice breaks (%)
Jitter parametrs	Jitter (local) (%)
	Jitter (local. absolute) (s)
	Jitter (rap) (%)
	Jitter (ppq5) (%)
	Jitter (ddp) (%)
Shimmer parameters	Shimmer (local) (%)
	Shimmer (local. dB) (dB)
	Shimmer (apq3) (%)
	Shimmer (apq5) (%)
	Shimmer (apq11) (%)
	Shimmer (dda) (%)
	Mean autocorrelation
Harmonicity parameters	Mean noise-to-harmonics ratio
	Mean harmonics-to-noise ratio
Frequence formant	F1 F2 F3 (Hz)
Intencity	Max inten Min inten mean inten

- Jitter (absolute) is the change in fundamental frequency F0 [25] from cycle to cycle, represented as [5]:

$$\text{Jitter (absolute)} = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i+1}|$$

- Jitter in Praat is a measure of how much the period (or pitch) of speech signals varies or perturbs. Praat offers different variants of jitter, such as Jitter (RAP), Jitter (PPQ5), and Jitter (ddp), which differ in the number of points used for calculation and the way they are expressed [11]. These variants provide different perspectives on the variation in pitch, allowing for more nuanced analysis of speech signals.

1.2.2. Shimmer measurement

Shimmer is a parameter that identifies micro-perturbations of the signal amplitude, and is measured using the parameters listed below [32]:

- Shimmer (relative) is used to describe a measure of the variation or amplitude perturbation in the intensity or loudness of speech signals. It is calculated as the relative change in amplitude between consecutive speech frames, expressed as a percentage [25].

$$\text{Shimmer (relative)} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} A_i}$$

where A_i represents the extracted peak-to-peak amplitude data, and N denotes the number of extracted fundamental frequency periods F0.

Shimmer (dB) is a speech and voice analysis measure that quantifies the amplitude or intensity variation in a vocal signal. It is expressed in decibels (dB) [11], it can be expressed as:

$$\text{Shimmer (dB)} = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20\log\left(\frac{A_{i+1}}{A_i}\right)\right|$$

- Shimmer, also referred to as Amplitude Perturbation Quotient (APQ), is a parameter used to assess the perturbation or fluctuation in the amplitude of speech signals. Various versions of shimmer exist, such as Shimmer (APQ3), Shimmer (APQ5), Shimmer (APQ11), and Shimmer (dda), which vary in terms of the number of data points utilized in the calculation and the manner in which they are expressed [11].

1.2.3. Harmonicity

The Mean Harmonics-to-Noise Ratio (HNR), which measures the proportion of harmonic components in the signal [7], and the Mean Noise-to-Harmonics Ratio (NHR), which evaluates the ratio of noise to harmonic energy in different frequency bands [8]. These measures are useful for analyzing the periodicity and noisiness of speech signals in various applications.

1.3. MFCC process

MFCCs (Mel-frequency cepstral coefficients) are a widely used feature extraction technique in audio processing and speech recognition. The Mel-scale is created by stacking triangular filters, usually ranging from 15 to 30, that are linearly spaced up to 1 kHz and logarithmically spaced above 1 kH [4]. The following stages are commonly included in computing MFCCs.

1.3.1. Farming

To account for the non-stationary character of speech waveforms over longer durations, a short-term analysis approach is utilized to efficiently analyze speech signals. Because the movement of speech articulators is physically restricted, frames of 10-30 ms are deemed stable for study. The analysis is carried out in regular time periods or frames, each with a fixed duration. This entails separating the voice signal into N-sample frames with an M-sample gap between consecutive frames (where M is fewer than N) [9].

1.3.2. Pre-emphasis

In this step, we amplify the energy in the voice stream by emphasizing higher frequencies. This is accomplished by solving the following first-order difference equation for the samples $\{s_n, n = 1, \dots, N\}$ [9]:

$$s'_n = s_n - k * s_{n-1}$$

The value of k in the equation represents the pre-emphasis coefficient, which is required to be within the range of $0 \leq k < 1$. according to reference [9]. For our investigation, we used a pre-emphasis coefficient of $k = 0.97$.

1.3.3. Fast Fourier Transform (FFT)

The audio frames that have undergone pre-processing are then subjected to the Fourier transform, which converts the signal from the time domain to the frequency domain. This process results in a representation of the audio signal in terms of its spectral content, which is defined for the set of N samples $\{s_n, n = 0,1,2, \dots, N-1\}$ as follow [4, 13]:

$$s_n = \sum_{k=0}^{N-1} s_k * e^{-2\pi jkn/N}$$

1.3.4. Mel-frequency wrapping

The Mel-filterbank is applied to the spectral representation of the audio signal using a set of triangular filters uniformly spaced in the Mel-scale, which mimics the frequency perception of the human ear. This filterbank is used to compute the energy in each Mel-frequency band. The Mel scale is logarithmic above 1000 Hz and linear below 1000 Hz, with a reference tone of 1 KHz at 40 dB above the perceptual hearing threshold set at 1000 mels. An approximate formula is used to determine the mels of a given frequency in Hz [13, 30].

$$Mel(f) = 2595 * log_{10}\left(1 + {}^{f}/_{700}\right)$$

1.3.5. Cepstrum

The compressed Mel-filterbank energies are subjected to a Discrete Cosine Transform (DCT), which serves to decorrelate the energies and capture cepstral features that represent the spectral shape of the audio signal. In order to transform the logarithm of the Mel spectrum into the time domain, the cepstral representation of the speech spectrum is utilized, achieved by applying the DCT to the log filter bank amplitudes using a specific formula. The MFCCs are valuable for frame analysis in speech processing, as they effectively capture the local spectral features of the speech signal [30].

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} m_j * \cos\left(\frac{\pi i}{N}(j - 0.5)\right)$$

The value of N corresponds to the total number of filter bank channels employed in the computation.

1.3.6. Liftering

Liftering is a technique in audio and speech signal processing that modifies MFCCs by applying a window function, such as a raised cosine window, to emphasize or attenuate specific frequency bands. It is commonly used to adjust spectral features in audio signals. One potential issue is that higher-order cepstral coefficients can become very small, which may pose challenges. To address this, cepstral liftering is employed, which involves normalizing the amplitudes of the cepstral coefficients using a specific formula [4, 9].

$$c'_n = (1 + \frac{L}{2} * \sin\left(\frac{\pi * n}{L}\right) * c_n$$

The value of L in the equation corresponds to the cepstral sine lifter parameter, and we employed $L = 22$ in this work.

2. Machine learning classifiers

After the feature selection process, the subsets of features are utilized in various machine learning algorithms to differentiate between PD patients and healthy individuals. We have employed three classifiers, namely Classification Tree (CT), Logistic Regression, Naive Bayes (NB), Support Vector Machines (SVM), and Ensemble methods.

The Classification Tree (CT) algorithm is an approach that uses decision tree principles to recursively split data based on feature values. It builds a tree structure where internal nodes represent features, and leaf nodes indicate class labels. The splits are determined by criteria like the Gini index or information gain, aimed at optimizing the separation between classes. This decision tree is a flowchart-like structure, where internal nodes symbolize functions or attributes, branches represent decision rules, and leaf nodes signify outcomes. The root node, situated at the top, decides the partitioning based on attribute values. Recursive partitioning is a technique for repeatedly dividing the tree. This flowchart-like representation aids in decision-making and closely mirrors human thinking processes. Consequently, decision trees are easily understandable and intuitive [15].

Logistic Regression is a popular mathematical method for predicting binary outcomes, where the class variable y takes values of 0 or 1. It differs from linear regression, which is more suitable for continuous outcomes, as logistic regression is specifically designed for categorical outcomes. It employs the standard logistic function, which is an S-shaped curve given by the equation:

$$f(x) = \frac{1}{1 + e^{-x}}$$

to calculate the probability of belonging to a particular class. The logistic regression model assumes a linear relationship between the input features and the logarithm of the odds of the outcome. By estimating the probability and applying a predefined threshold, the algorithm assigns the appropriate class label to each instance [15].

Naive Bayes (NB) is a probabilistic classifier that applies Bayes' theorem with the assumption of feature independence. It calculates the probability of each class given the input features and assigns the class label with the highest probability. Despite its "naive" assumption, Naive Bayes has shown good performance in many classification tasks [15].

Naive Bayes classifier is a probabilistic machine learning model that is used for binary (two-class) and multi-class classification problems. The classifier is based on the Bayes theorem:

$$P(y|X) = \frac{P((X|y)P(y)}{P(X)}$$

The class variable y and the parameter/feature variables represented by $X = (x1, x2, \dots, xn)$ are involved in Naive Bayes. The Naive Bayes classifier assumes that the attributes are independent of each other. Hence, it assumes that the presence or value of one attribute does not affect the presence or value of other attributes.

$$P(y|X) = \frac{P(x_1|y)P(x_2|y) \dots P(x_n|y)P(y)}{P(x_1)P(x_2) \dots P(x_n)}$$

For Gaussian Naive Bayes, the conditional probability is derived from a normal distribution, similar to a Gaussian distribution.

$$P(x_i|y) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-{(x_i-\mu_y)^2}/{2\sigma^2 y}}$$

Support Vector Machines (SVM) is a supervised machine learning technique utilized for solving classification and regression problems. It distinguishes itself by identifying a hyperplane that effectively separates the classes. The hyperplane is determined by maximizing the margin, which is the distance between the hyperplane and the nearest data points from each class. In cases where the data is not linearly separable, the kernel trick is employed. The kernel function allows for transforming the input space from a lower dimension to a higher dimension, enabling the resolution of non-linear separable problems. In our research, we utilized the sigmoid function as the kernel for SVM [15].

Ensemble methods combine multiple individual classifiers to make predictions. Techniques such as Random Forest, Bagging, or Boosting can be used for ensemble methods. These methods create an ensemble of classifiers and aggregate their predictions to make a final decision. Ensemble methods often improve classification accuracy and help mitigate overfitting.

3. Evaluation metrics

In order to evaluate the effectiveness of our classifiers in distinguishing between PD patients and patients with different ND, we used many performance criteria, including accuracy, sensitivity, and specificity, to assess the usefulness of our classifiers in differentiating between PD patients and patients with other ND [4, 6].

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

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

True positives (TP) are Parkinson's disease patients successfully diagnosed by the classifier. True negatives (TN) are patients with ND who have been appropriately classified. False positives (FP) are ND patients who were wrongly categorized. False negatives (FN) are Parkinson's disease individuals who were misclassified [6, 30]. Accuracy quantifies the classifier's performance in discriminating between the two groups, sensitivity measures the accuracy of PD patient recognition, and specificity gauges the accuracy of detecting patients with other neurodegenerative diseases [6, 9]. Additionally, we included two more variables along with accuracy, sensitivity, and specificity. Matthews' correlation coefficient (MCC) and Probability Excess (PE). MCC is a reliable measure of the quality of binary classification and ranges from 0 (for random algorithms) to 1 (for perfect algorithms). PE, on the other hand, also ranges from 0 (for random prediction) to 1 (for perfect prediction) [4].

$$MCC = \frac{TP \times TN - FN \times FP}{\sqrt{((FN + TP)(FP + TN)(FP + TP)(FN + TN))}}$$

$$PE = \frac{TP \times TN - FN \times FP}{(FN + TP)(FP + TN)}$$

Table 2. Confusion matrix used in the analysis

		Result	Of classification
		ND	PD
Diagnosed	ND	TP	FP
	PD	FN	TN
		Sensitivity	Specificity

4. Results and discussion

In our research, we utilized cepstral analysis to extract features and coefficients from voice samples in order to classify patients into two categories: those with Parkinson's Disease (PD) and those without (ND). We employed various classifiers and kernels, such as Classification Tree (CT), Logistic Regression, Naive Bayes (NB), Support Vector Machines (SVM), and Ensemble methods, to accurately differentiate between the two groups and gain insights into the potential use of voice analysis in PD diagnosis. To address the issue of overfitting, we con-ducted cross-validation experiments with different numbers of folds, including 5, 10, 20, and 30, on our dataset which com-prised 76 samples with 75 features each. After careful analysis of the results, we determined that using 20 folds for cross-validation provided accurate performance estimation.

The table 3 presents the classification performance of various acoustic features for differentiating Parkinson's Disease (PD) from other neurodegenerative diseases (ND) without employing dimensionality reduction techniques such as PCA. The results demonstrate the accuracy, sensitivity, specificity, true positive (TP), true negative (TN), false positive (FP), false negative (FN), Matthewscorrelation coefficient (MCC), and classification error rate (PE) for each feature. Overall, the table shows that some features achieved high accuracy of 73.7% or 75% in differentiating PD from ND. However, it is important to note that several features, such as median pitch, mean pitch, standard deviation of pitch, minimum pitch, maximum pitch, mean period, standard deviation of period, jitter parameters, shimmer parameters, mean autocorrelation, and harmonicity parameters, all had an accuracy of 73.7% without distinguishing between PD and ND. These features did not provide sufficient discriminatory power for accurate classification. On the other hand, certain features exhibited higher accuracy and sensitivity values, such as the number of pulses (77.63%), number of periods (78.94%), fraction of locally unvoiced frames (73.68%), number of voice breaks (75%), degree of voice breaks (77.63%), jitter (ppq5) (75%), formant frequencies (F1, F2, F3) (77.5%), and the first 15 coefficients of MFCC (75%). These features showed promise in differentiating PD from other ND with reasonable accuracy. However, it is crucial to consider the limitations of the study and the interpretation of these results. The sample size used in this study was relatively small, with only 56 PD patients and 20 patients with other ND. Therefore, the generalizability of the findings to a larger population should be carefully considered. Additionally, the specific classification algorithms and parameter settings used for the machine learning models are not provided in the table, making it difficult to assess the reproducibility and reliability of the results.

In conclusion, the table highlights the classification performance of different acoustic features in distinguishing PD from other ND without employing dimensionality reduction techniques. Some features exhibited higher accuracy and sensitivity, while others showed limited discriminatory power. Further research with larger and more diverse datasets, along with detailed information on classification algorithms and parameter settings, is necessary to validate and improve the effectiveness of these features for accurate PD classification.

The performance of various classifiers evaluated using linear PCA can be assessed by selecting the best 4 numbers from the 13th principal components. In Table 4, it was observed that Kernel NB, with numeric components 2, 4, 6, and 7, consistently demonstrated superior performance in terms of accuracy (ranging from 80.3% to 86.84%), sensitivity (ranging from 72.77% to 89.65%), specificity (ranging from 77.77% to 89.65%), and MCC (ranging from 0.5837 to 0.6743). Gaussian NB also showed relatively good performance with accuracy ranging from 75% to 78.94% and sensitivity ranging from 50% to 68.75%. The number of numeric components used in linear PCA appeared to impact the performance of certain classifiers. SVM and ensemble (Bugged trees) showed consistent accuracy of 73.7% across all numeric component values, indicating that the performance was not influenced by the number of components. However, Rus Boosted trees displayed a significant drop in accuracy from 34.21% to 28.57% as the number of components increased from 2 to 4. Sensitivity, also known as recall or true positive rate, measures the ability of a classifier to accurately identify positive cases, while specificity, also known as true negative rate, measures the ability to correctly identify negative cases. Higher sensitivity and specificity values are generally desired for a reliable classifier, as they indicate better capability in correctly classifying both positive and negative cases. MCC is a metric that assesses the quality of binary (two-class) classification, taking into account true positives, true negatives, false positives, and false negatives. A higher MCC value indicates better classification performance. PE measures the average prediction error of the classifier, with lower values indicating better performance.

Table 3. Classification Performance of Acoustic Features for Differentiating Parkinson's Disease and Other Neurodegenerative Diseases without Dimensionality Reduction

Groups	Features	Accuracy (%)	Sensitivity (%)	Specificity (%)	TP	TN	FP	FN	MCC	PE
Pitch parameters	Median pitch	73.7	0	100	0	56	20	0	0	0
	Mean pitch	73.7	0	100	0	56	20	0	0	0
	Standard deviation	73.7	0	100	0	56	20	0	0	0
	Minimum pitch	73.7	0	100	0	56	20	0	0	0
	Maximum pitch	73.7	0	100	0	56	20	0	0	0
Pulses parameters	Number of pulses	77.63	63.63	80	7	52	13	4	0.3486	0.2823
	Number of periods	78.94	63.63	80.30	7	53	13	3	0.3891	0.4643
	Mean period	73.7	0	100	0	56	20	0	0	0
	Standard deviation of period	73.7	0	100	0	56	20	0	0	0
Voicing parameters	Fraction of locally unvoiced frames	73.68	5.3	75.71	3	53	17	3	0.1574	0.2571
	Number of voice breaks	75	10	75.34	2	55	18	1	0.1857	0.42
	Degree of voice breaks	77.63	80	77.46	4	55	16	1	0.3235	0.5746
Jitter parametrs	Jitter (local)	73.7	0	100	0	56	20	0	0	0
	Jitter (local. absolute)	73.7	0	100	0	56	20	0	0	0
	Jitter (rap)	73.7	0	100	0	56	20	0	0	0
	Jitter (ppq5)	75	3.44	75.67	2	56	18	0	0.275	0.7567
	Jitter (ddp)	73.7	0	100	0	56	20	0	0	0
Shimmer parameters	Shimmer (local)	73.7	0	100	0	56	20	0	0	0
	Shimmer (local. dB)	73.7	0	100	0	56	20	0	0	0
	Shimmer (apq3)	73.7	0	100	0	56	20	0	0	0
	Shimmer (apq5)	73.7	0	100	0	56	20	0	0	0
	Shimmer (apq11)	73.7	0	100	0	56	20	0	0	0
	Shimmer (dda)	73.7	0	100	0	56	20	0	0	0
Harmonicity parameters	Mean autocorrelation	73.7	0	100	0	56	20	0	0	0
	Mean noise-to-harmonics ratio	73.7	0	100	0	56	20	0	0	0
	Mean harmonics-to-noise ratio	73.7	0	100	0	56	20	0	0	0
Frequence formant	F1 F2 F3	77.5	60	72.22	6	52	14	4	0.2977	0.3878
Intencity	Max inten, Min inten, mean inten	73.7	0	100	0	56	20	0	0	0
Coefficient MFCC	15 firsts Coeff	75	3.44	75.67	2	52	18	0	0.275	0.7567

Table 4. Performance Comparison of Different Classifiers with Varying Number of Numeric Components

Classifiers		Number numeric of components	Acc (%)	Sens (%)	Spec (%)	TP	TN	FP	FN	MCC	PE
CT		2 4 6 7	73.7	0	100	0	56	20	0	0	0
LR		2 4 6 7	73.7	0	100	0	56	20	0	0	0
Navies Bayes	Gaussian NB	2	75	53.85	79.37	7	50	13	6	0.2839	0.3321
		4	78.9	68.75	0.845	11	49	9	5	0.4934	0.5323
		6	75	53.85	79.37	7	50	13	6	0.2839	0.3289
		7	78.94	66.66	86.66	8	52	12	4	0.3968	0.4792
	Kernel NB	2	80.3	77.77	80.59	7	54	13	2	0.4283	0.5837
		4	82.9	73.33	85.24	11	52	9	4	0.5295	0.5858
		6	84.21	72.77	87.93	13	51	7	5	0.5808	0.6015
		7	86.84	77.77	89.65	14	52	6	4	0.6510	0.6510
SVM		2 4 6 7	73.7	0	100	0	56	20	0	0	0
Ensemble	Bugged trees	2 4 6 7	73.7	0	100	0	56	20	0	0	0
	Rus Boosted trees	2 4 6 7	34.21	28.57	100	20	6	56	0	0.6732	0.2857

In summary, after analyzing the outcomes of linear PCA on various classifiers, it can be concluded that Kernel NB consistently demonstrates superior performance across different numeric component values, followed by Gaussian NB. However, the choice of the best classifier may also depend on specific requirements and goals of the classification task, and other factors like computational efficiency, interpretability, and ease of implementation should also be taken into consideration. It is crucial to thoroughly evaluate the performance of different classifiers using appropriate evaluation metrics and select the one that aligns with the specific needs of the task at hand.

5. Conclusion

The study aimed to assess speech impairments in patients with Parkinson's disease and other neurodegenerative diseases by anlyzing acoustic measurements over time, as well as spectral and cepstral measurements. The researchers extracted acoustic features and MFCCs from voice samples using cepstral domains, and then performed feature selection through PCA to differentiate between PD patients and those with other neurodegenerative diseases. The results showed that using PCA with the Naive Bayes kernel and linear classification, the maximum classification accuracy achieved was 86.84%. This study demonstrated the potential of using objective assessments and machine learning techniques to accurately identify and differentiate speech disorders in PD patients, which could lead to improved management and treatment of these disorders, ultimately enhancing communication and quality of life for patients. The researchers utilized various classifiers and kernels, along with cross-validation, to address overfitting and classify patients with PD and other neurodegenerative diseases based on voice samples. The Naive Bayes kernel with PCA yielded the most favorable results in terms of classification accuracy. However, when selecting the best classifier, specific task requirements and other factors such as computational efficiency, interpretability, and ease of implementation should be considered. Thorough evaluation of classifier performance using appropriate metrics is crucial in choosing the most suitable classifier for the task at hand.

Future work should consider incorporating deep learning and convolutional neural networks (CNN) to assess the severity of the disease. This can improve the accuracy and specificity of speech disorder classification in Parkinson's disease (PD) and other neurodegenerative diseases. Furthermore, the inclusion of larger datasets and exploration of alternative machine learning algorithms can provide valuable insights and enhance the overall performance of classification models. These advancements have the potential to enhance our understanding of speech impairments in neurodegenerative diseases and contribute to personalized treatment approaches.

References

[1] Abramson E. L. et al.: Physician experiences transitioning between an older versus newer electronic health record for electronic prescribing. International journal of medical informatics 81(8), 2012, 539–548.
[2] Andersen T. et al.: Designing for collaborative interpretation in telemonitoring: Re-introducing patients as diagnostic agents. International journal of medical informatics 80(8), 2011, e112–e126.
[3] Baker K. K. et al.: Thyroarytenoid muscle activity associated with hypophonia in Parkinson disease and aging. Neurology 51(6), 1998, 1592–1598.
[4] Benba A. et al.: Discriminating between patients with Parkinson's and neurological diseases using cepstral analysis. IEEE transactions on neural systems and rehabilitation engineering 24(10), 2016, 1100–1108.
[5] Benba A. et al.: Hybridization of best acoustic cues for detecting persons with Parkinson's disease. Second World Conference on Complex Systems (WCCS), IEEE, 2014, 622–625.
[6] Benba A. et al.: Using RASTA-PLP for discriminating between different neurological diseases. International Conference on Electrical and Information Technologies (ICEIT), IEEE, 2016, 406–409.
[7] Boersma P. et al.: Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proceedings of the institute of phonetic sciences, 1993, 97–110.
[8] De Colle W.: Voce & Computer – analisi acustica digitale del segnale verbale. Omega, 2001.
[9] Gales M. et al.: The application of hidden Markov models in speech recognition. Foundations and Trends in Signal Processing 1(3), 2008, 195–304.
[10] Gillies G. E., et al.: Sex differences in Parkinson's disease. Frontiers in neuroendocrinology 35(3), 2014, 370–384.
[11] Gorris C. et al.: Acoustic analysis of normal voice patterns in Italian adults by using Praat. Journal of Voice 34(6), 2020, 961.e9–961.e18.
[12] Goyal J. et al.: A hybrid approach for Parkinson's disease diagnosis with resonance and time-frequency based features from speech signals. Expert Systems with Applications 182, 2021, 115283.
[13] Gupta S. et al.: Feature extraction using MFCC. Signal & Image Processing. An International Journal 4(4), 2013, 101–108.
[14] Holi M. S. et al.: Automatic detection of neurological disordered voices using mel cepstral coefficients and neural networks. IEEE Point-of-Care Healthcare Technologies (PHT), IEEE, 2013, 76–79.
[15] Islam M. R., Nahiduzzaman M.: Complex features extraction with deep learning model for the detection of COVID19 from CT scan images using ensemble based machine learning approach. Expert Systems with Applications 195, 2022, 116554.
[16] Lieberman P.: Perturbations in vocal pitch. The Journal of the Acoustical Society of America 33, 1961, 597–602.
[17] Lieberman P.: Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. The Journal of the Acoustical Society of America 35(3), 1963, 344–353.
[18] Little M. et al. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. Nature Precedings, 2008, 1–1.
[19] Little M. et al.: Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Nature Precedings, 2007, 1–1.
[20] Mandal I., Sairam N.: Accurate telemonitoring of Parkinson's disease diagnosis using robust inference system. International journal of medical informatics 82(5), 2013, 359–377.
[21] O'Sullivan S. B., Schmitz T. J.: Parkinson disease. Physical Rehabilitation 5th ed., F. A. Davis Company, Philadelphia 2007, 856–894.
[22] Pahuja G., Nagabhushan T. N.: A comparative study of existing machine learning approaches for Parkinson's disease detection. IETE Journal of Research 67(1), 2021, 4–14.
[23] Parkinson J.: An Essay on Shaking Palsy. Whittingham and Rowland Printing, London 1817.
[24] Rahman A. et al.: Parkinson's disease diagnosis in cepstral domain using MFCC and dimensionality reduction with SVM classifier. Mobile Information Systems 2021, 1–10.
[25] Rahn III. et al.: Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis. Journal of Voice 21(1), 2007, 64–71.
[26] Sharanyaa S. et al.: An Exploration on Feature Extraction and Classification Techniques for Dysphonic Speech Disorder in Parkinson's Disease. Inventive Communication and Computational Technologies: Proceedings of ICICCT 2021, Springer Singapore, 2022, 33–48.
[27] Stelzig Y. et al.: Laryngeal manifestations in patients with Parkinson disease. Laryngo-rhino-otologie 78(10), 1999, 544–551.
[28] Teixeira J. P. et al.: Vocal acoustic analysis–jitter, shimmer and hnr parameters. Procedia Technology 9, 2013, 1112–1122.
[29] Teston B.: L'évaluation objective des dysfonctionnements de la voix et de la parole; 2e partie: les dysphonies. Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA) 20, 2001, 169–232.
[30] Tiwari V.: MFCC and its applications in speaker recognition. International journal on emerging technologies 1(1), 2010, 19–22.
[31] Van Den E. et al.: Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. American Journal of Epidemiology 157(11), 2003, 1015–1022.
[32] Wendahl R. W.: Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. Folia Phoniatrica et Logopaedica 18(2), 1966, 98–108.

M.Sc. Oumaima Majdoubi
e-mail: oumaima_majdoubi@um5.ac.ma

Received the master's degree in Electrical Engineering from the National School of Arts and Crafts of Rabat, Mohammed V University, Rabat, Morocco, in 2022. She is a research student of sciences and Technology of the Engineer in Ecole Nationale Supériure d'Informatique et d'Analyse des Systémes ENSIAS. She is part of E2SN research team at ENSAM-RABAT. She is interested in the detection of neurological diseases based on artificial intelligence and signal processing.

http://orcid.org/0009-0000-2968-7975

Prof. Achraf Benba
e-mail: achraf.benba@ensam.um5.ac.ma

Received the Ph.D. degree in electrical engineering from ENSAM, Mohammed V University, Rabat, Morocco, in 2017. He is a member of E2SN teams at ENSAM, Mohammed V University. His interests are in speech processing for detecting people with neurological disorders and Cardiac.

http://orcid.org/0000-0001-7939-0790

Prof. Ahmed Hammouch
e-mail: ah.hammouch@gmail.com

Received the M.S. degree and the Ph.D. degree in automatic, electrical, electronic by the Haute Alsace University, Mulhouse, France, in 1993, and the Ph.D. degree in signal and image processing by the Mohammed V University, Rabat, Morocco, in 2004. From 1993 to 2013, he was a professor in the Mohammed V University, Rabat, Morocco. Since 2009 he manages the Research Laboratory in Electronic Engineering. He is an author of several papers in international journals and conferences. His domains of interest include multimedia data processing and telecommunications. He is with National Center for Scientific and Technical Research, Rabat, Morocco.

http://orcid.org/0009-0005-8691-6662