Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 1 (2024)

Investigation of Acoustic Features and Machine Learning for Early Detection of Parkinson's Disease

Hadjer Zebidi[1], Dr. Zeineb Benmasaaoud[2], Pr. Mondher Frikha[3]
[1,3] National School of Electronics and Telecommunications (ENET'Com), University of Sfax, Tunisia
[2] Higher Institute of Computer Science and Multimedia of Gabès (ISIMG), Tunisia

Abstract:- Parkinson's disease (PD) is a neurodegenerative disease that afflicts millions of people. The early detection of the disease is crucial. According to recent research, the level of dysarthria is a good indicator for computer-assisted diagnosis and remote monitoring of patients in the early phases. Despite the significance of articulatory deficits in dysarthria among individuals with PD, automatic speech performance evaluation methods mainly concentrate on assessing dysphonia. In this study, our objective was to classify the phonation, articulation et diadochokinetic features by machine learning (ML) algorithms with feature selection technique. Using Italian data, a Lasso-cross validation feature selection algorithm was used to select the voice features extracted from three vocal tasks, followed by three classifiers (Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Machine (SVM)) to detect the disorder. The highest overall classification score achieved an 100% accuracy rate in discriminating between PD and control participants. More interest, articulatory features were found to be the most powerful indicator of PD-related dysarthria than phonation and DDK features among all the classifying algorithms.

Keywords: *Parkinson's disease, early detection, voice features, Machine Learning, Lasso-Cross Validation, Artificial Neural Network, Random Forest, Support Vector Machine*

1. Introduction

Parkinson's disease is a neurodegenerative condition, ranked as a second most prevalent after Alzheimer's disease. It impacts the central nervous system with the gradual loss specific neurons cells within the brain. The primary reason for this degeneration is the deterioration of dopamine produced neurons located within the substantia nigra, the specific area of the brain responsible for movement control. This condition is typically diagnosed in individuals aged over 65 years [1]. Motor symptoms such as tremors, slowness of movement, rigidity of the body, and balance problems are the primary manifestations of this disease. Advanced neuroimaging technologies like magnetic resonance imaging (MRI) have led to a better comprehension of the patients' neuropathological processes. Nevertheless, the diagnostic procedure is intricate, requiring a specialist physician, and is usually conducted during advanced stages of the disease, taking a prolonged period of time. Most studies currently rely on MRI imaging as the primary and most relevant clinical diagnostic tool. However, its expense and invasiveness restrict its use as a screening method, and researchers are therefore exploring biomarkers present in more accessible areas of the body. Diagnosis can be challenging in cases where symptoms are incomplete or subtle, and as a result, some individuals may remain undiagnosed.

This challenge motivates the creation of computer-based tools, decision aids and tests that can support early diagnosis and predict the progression of these diseases. One possible early indication of disease onset is vocal deterioration. Modification of the patient's voice is among the multiple clinical presentations of these diseases and appears to be a noteworthy element for various reasons. Modification of the patient's voice is among the multiple clinical presentations of these diseases and appears to be a noteworthy element for various reasons. Numerous studies have analyzed voice disorders in Parkinson's patients, revealing that 90% of individuals with PD suffer from associated speech difficulties [2]. The most frequently observed symptoms and disturbances include reduced voice volume, monotony of voice, changes in the voice's quality, rate of speech, and uncontrollable repetition of words. Voice production is a complex process that involves various speech subsystems working together to create speech sounds. These subsystems include the respiratory system (lungs, trachea), phonatory system (larynx, vocal cords), and articulatory system (pharynx, jaw, lips, tongue, nasal cavities, palate, velum). Additionally, breathing plays a crucial role in speech production and requires additional coordination. For an individual with this condition, it can be challenging to measure and modify the extent of movements required to execute a precise task, including voice production.

The paper is organized into several sections: some related works to the classification of Parkinson's disease are mentioned in section II. In section III, the materials and methods presented in the study are analyzed, it's elaborated the details regarding the database used and discusses the early detection model proposed for Parkinson's disease, presenting subsections on the extraction of features, selection of relevant features, with classification methods. Section IV incorporates results, provides a comprehensive account of the experimental outcomes and makes a discussion about the results. the paper is providing finally some concluding remarks.

2. Related Work

This section provides an overview of various studies on the detection and diagnosis of Parkinson's disease through speech. Quan & al. [3] proposed the use of a bidirectional long-short term memory (Bi-LSTM) model to capture the dynamic features of the time series of a speech signal in order to detect Parkinson's disease. During the same year, researchers including Harel et al., Skodda and Visser [3] studied sustained vowel /i/, short sentence repetition, and monologue text reading in 90 seconds. They assessed these factors by measuring articulation and phonation functionalities. The disease was classified using both linear and non-linear classifiers such as SVM, ANN, and DNN.

In addition, Fagherazzi & al. [4] analyzed an individual's voice using artificial intelligence (AI) to identify and use vocal biomarkers. The vocal signal was pre-processed, linguistic and acoustic features were extracted through MFCC to diagnose, predict the risk of, and remotely monitor various clinical outcomes and disease symptoms.

J. R. Orozco-Arroyave & al. [5] conducted a speech analysis on individuals with neurodegenerative conditions, with a focus on patients with Parkinson's Disease. The study used objective measurements to assess patients' communication abilities, specifically evaluating voice impairment in terms of phonation, articulation, prosody, and intelligibility. The assessment measures biomarkers linked to dysarthria levels in PD patients and predicts disease progression via two assessment scales. The Unified Parkinson's Disease Rating Scale (UPDRS) evaluates behavior, mood, activities of daily living, motor tasks and therapeutic effect, while the Hoehn & Yahr Scale exclusively scores gait and posture disorders.

In 2022, P.Fan [6] is studied the effectiveness of identifying PD patients from speech signals by various acoustic parameters including prosodic and segmental features are extracted from speech, and then RF classification algorithm based on these acoustic parameters is applied to diagnose early disease.

Several studies have utilized AI techniques to analyze speech features. Vergara et al. [3] employed models based on ML including SVM, DT, and Multi-Layer Perceptrons (MLP). In addition, they implemented Deep Learning models like Convolutional Neural Networks (CNN) and Recurrent Neural Networks. Their method involved feature extraction including articulation and prosody. They used classifiers with reading of sustained vowel /a/ and repetition of short sentence to deduce whether a patient has PD or is healthy. To assess the performance, they measured Accuracy, F-score, and Matthews correlation coefficient (MCC).

Using a novel speech analysis technique adapted from the recent advances in speaker recognition, L. Jeancolas et al. [7] investigated the differentiation between individuals in the early stages of Parkinson's disease and those who are healthy. They utilized new quantitative biomarkers, referred to as "x-vectors", to scrutinize the influence of these parameters (audio segment duration, data argumentation, type of dataset applied in neural network training, type of speech task, and back-end analysis) on the performance of disease classification. This approach involves extracting integrations from deep neural networks, using MCC as input. Technical term abbreviations are explained upon first use. Spelling follows British conventions. It results in robust speaker representation and improved patient recognition when more test data is used for training. Additionally, efficiency and accuracy of the latest classification technique (X-vector) was compared to the classic MFCC-GMM technique, revealing the effect of gender in classification. The language is clear, objective, formal, avoiding biased or emotional language, with clear structure, and causal connections between statements. According to the study, x-vector is a more effective classifier in detecting Parkinson's disease in women, with enhancements between 7% and 15% in comparison to other traditional techniques.

A further study aimed to detect Parkinson's disease early for the sake of earlier treatment. Kumari & Ramachandran [8] analyzed speech signals using ML methods to distinguish the disease. They utilized MFCC and LPC parameters to describe the timbre and various classification algorithms such as ANN and CNN. Finally, it was concluded that these models perform better in terms of maximum accuracy but have certain limitations. Classification algorithms based on ML and DL such as ANN and CNN were used.

The research aimed to develop an automatic Chinese data detection method. Wang et al. [9] extracted the acoustic characteristics of both phonation and articulation. They employed three selection algorithms - Lasso, mRMR, and Relief-f to determine the most representative features and the most accurate algorithm for selection. The Lasso algorithm outperformed the others. The subsequent step involved the application of four ML classifiers to identify speech impairment in PD patients. The Logistic Regression (LR) classifier model achieved the highest sensitivity at 82.44%, while SGD yielded an accuracy of 75.76%. According to this study, articulation features (F1, F2, BBE, MFCC) were found to be more indicative and representative than phonation parameters across all selections and classifications.

S. Dheer et al. [10] discovered that the incorporation of phonation and articulation into extracted features can attain acceptable outcomes in identifying patients with Parkinson's disease (PD). Additionally, various features generally utilized in speech processing for predicting and assessing disease progression, such as the UPDRS scale, were also employed in detecting Parkinson's disease. Using data extracted from the UPDRS and parameters, ML-based algorithms such as K-Nearest Neighbor (KNN), **Linear Discriminant analysis** (LDA), **Decision Tree** (DT), **Neural Network** (NN), and Gradient Boosting (GB) were employed to create a classification framework consisting of 24 speech features. Technical abbreviations have been explained when first used. The KNN classifier was found to be the most effective model for detecting PD, with an accuracy score of 97.96% and an MCC of 0.93674, demonstrating a high level of agreement between the input and the trained model.

The approach of studying acoustic waves for automatic and early detection of Parkinson's disease is discussed in several papers. M. Alalayah et al. [11] focused on the classification algorithms based on the recursive feature elimination (RFE) method to identify the disease. The identification of Parkinson's disease at an early stage can prove to be challenging. The present study aims to tackle the difficulties linked with the early detection of Parkinson's disease as it can significantly improve patient outcomes. The researchers propose that examining acoustic waves could provide a non-invasive and cost-effective approach to disease detection. Multiple acoustic properties were extracted from databases. The RFE method was utilized to determine the most relevant features for selection, alongside three classification techniques (SVM, KNN, RF). The study demonstrated that RFE-based classification algorithms achieved a success rate of 97% to 98% in accurately identifying Parkinson's disease. The researchers concluded that their approach holds great potential for the early and automatic identification of Parkinson's disease via acoustic signals. They suggest that to facilitate universal screening and early intervention, their method could be enhanced and integrated into a smartphone application.

3. Materials And Methods

A. Proposed detection architecture

In this study, we are proposed a model for detecting and differentiating patients with Parkinson's disease from healthy controls. This model includes Italian database, feature extraction, feature selection, and classification models, as shown in Fig. 1.

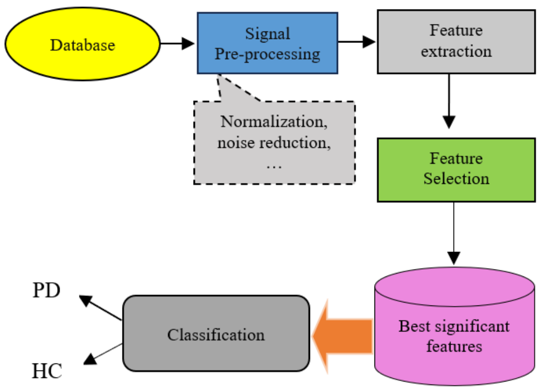


Fig. 1. Overview of the proposed detection architecture

B. Database

The database utilized in the research was delivered by Multimedia Electronic Health Record and developed by Giovanni and other authors [13]. Our study collected voice recordings from 47 Italian speakers in the early stage, all patients were rated as stage 1 to 2 on the UPDRS assessment scale including 25 PD patients (17 male & 8 female) and 22 healthy controls (10 male & 12 female) [14]. Patients ranged in age from 40 to 80 years. Voice recordings were made in a quiet and echo-free environment while keeping the microphone at a distance of 15-25 cm from the mouth. This protocol necessitated speech production in different modes and execution times.

Clinical practice has shown that sustained vowels [12] and running speech are good materials for detection. This study uses three voice tasks for speech analysis: sustained vowels, isolated words, and repetition of syllable.

We were collected for the dataset, the sustained vowel /a/ uses in analysis of phonatory while the word test was added for articulation in the continuous speech analysis and repetition of the syllable ‘ta’ for Diadochokinetic (DDK) analysis. A total of 50 balanced words were recorded for each subject in the database and They are a common word of our daily lives.

C. Features extraction

There is an extensive body of literature on voice studies in Parkinson's disease, which illuminates disorders such as hypokinetic dysarthria. This is characterized by reduced prosody, difficulties with articulation, irregular phonation, cognitive, and memory impairments. There are several acoustic features related to PD detection in the early stage. In this paper, we analyzed the different voice disorders that occur in Parkinson's patients, including phonation, articulation, and Diadochokinetic.

Phonatory analysis

Phonation is characterized by insufficient curvature and closure of the vocal cords. In our study, we were evaluated through a series of perturbation measures extracted such as jitter (Temporal disturbance of the frequency of the audio signal), shimmer (Temporal perturbation of the amplitude of the audio utterance), Amplitude Perturbation Quotient (APQ), Pitch Perturbation Quotient (PPQ), and Degree Unvoiced.

Jitter was obtained in (1):

$$\text{Jitter (\%)} = \frac{100}{N \cdot Mf} \sum_{k=1}^{N} |F_0(k)-Mf| \quad (1)$$

Where: N: Number of frames of the speech utterance, Mf: Maximum of F0 and F0(k) corresponds to the F0 calculated on the k-th frame.

Shimmer was calculated as shown in (2):

$$\text{Shimmer (\%)} = \frac{100}{N \cdot Ma} \sum_{k=1}^{N} |A(k)-Ma| \quad (2)$$

Where N is a number of frames of the speech utterance, Ma is maximum amplitude of the signal, and A(k) corresponds to the amplitude on the k-th frame.

APQ were computed as measures of the variability of the amplitude and pitch of the peak-to-peak voice signal with a smoothing factor of 11 periods, PPQ measures the variability of F0 with a smoothing factor of 5 periods.

The two perturbation quotients (PQ) are calculated using the equation below (3).

$$PQ = \frac{1}{L}\sum_{i=1}^{L}\frac{\frac{1}{k}\sum_{j=1}^{k} S(i+j-1)-S(i+n)}{\left|\frac{1}{M}\sum_{j=1}^{M} S(i)\right|} \quad (3)$$

Where: L = N-(k-1), n = (k-1)/2

S: The pitch period sequence (PPS) when calculating the PPQ, the Pitch Amplitude Sequence PAS when calculating the APQ

N: The length of the PPS or PAS

k: The length of the moving average (11 for PAQ or 5 for PPQ)

Degree Unvoiced is percentage of the sustained vowel utterance /A/ that is detected as unvoiced using voiced-unvoiced frame segmentation [5].

Articulatory analysis

Articulation is related to the reduced amplitude and speed of movement of the lips, jaws, and mandible. Concerning continuous speech signals, articulation features are measured by calculating:

The spectrum of the transitions is distributed into 22 critical bands following the Bark scale, and the Bark-band energies (BBE) are calculated according to the Bark scale. We are evaluated by computing the energy content of the transition from unvoiced to voiced segments (onset) and the transition from voiced to unvoiced segments (offset) to assess the patient's ability to control the muscles that produce the voice, such as the tongue.

Formula (4) reproduces the frequency distribution suggested by Zwicker et al. [15], that the Bark scale function can be expressed analytically as:

$$Bark(f) = 13\ \tan^{-1}\left(0.76\frac{f(HZ)}{1000}\right) + 3.5\tan^{-1}\left(\frac{f(HZ)}{7500}\right)^2 \quad (4)$$

Mel-frequency cepstral coefficients (MFCC) are the most frequently used to characterize the spectral envelope which reflects the shape of the vocal cords. We extracted MFCC using the Librosa library with a 22050 sample rate. The feature vectors were then enriched by computing the first (12 deltas MFCC) and second (12 delta-deltas MFCC) derivatives to provide additional speech dynamic information.

The formula for converting from frequency to Mel scale [16] is shown in (5):

$$M(f) = 1127\ln\left(1+\frac{f}{700}\right) \quad (5)$$

where f is the Hz frequency

Analysis Diadochokinetic

DDK examine the patient's capacity to manipulate their articulators, specifically their jaw and tongue movements. It assessed through measurements related to speech rate, speech duration, speech frequency F0, and speech energy that are:

- The variability of the fundamental frequency measured in semitones.
- The variability of the fundamental frequency measured in Hz.
- the average energy of the DDK task.
- The variability of the energy along the utterance.
- The maximum value of the energy.
- The DDK rate, that is measured number of syllables are uttered by second.
- The DDK regularity, which is measured as the variability of the duration of each syllable.
- The average duration of the syllables.
- The pause rate that is measured number silence segments appear per second.

- The average duration of pauses.
- The regularity of pauses, which is measured as the variability of the duration of the pauses.

In our research, we extracted phonation, articulation, and DDK features that imported them into the extraction features algorithm as a CSV file. In the dataset, the "class" column is designated as 0 for HC and 1 for PD to distinguish healthy individuals from those with PD.

D. Features selection

In this phase, Lasso-CV (Least Absolute Shrinkage and Selection Operator Cross Validation) was applied to select the most important features from a dataset and for dimensionality reductions of the dataset. Lasso-CV implements a more developed and powerful approach to feature selection, offering improved model stability, automatic selection of the regularization parameter and potentially better predictive performance. This means that Lasso-CV can automatically select the most relevant features and assign zero coefficients to the irrelevant or redundant ones. This is a method for selecting features that combine Lasso regression with cross-validation. It employs K-fold cross-validation to determine the optimal value of the regularization parameter (alpha) that establishes the amount of shrinkage applied to the coefficients, facilitating the selection of the most significant features from a given dataset [17]. This study is employed under method 10-fold cross validation (CV) to find the best alpha that minimizes the mean squared error (MSE), on Parkinson disease dataset using Scikit-learn.

The objective of Lasso-CV is to identify the most optimal α that minimizes the cost function outlined in (6).

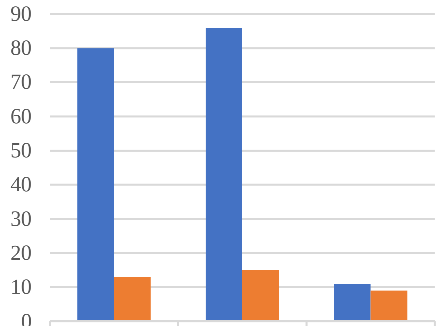
$$min_{\alpha}\left[\frac{1}{2n}\sum_{i=1}^{n}\left(y_i-\left(\beta_0+\sum_{j=1}^{p}\beta_j\,x_{ij}\right)\right)^2+\alpha\sum_{j=1}^{p}\left|\beta_j\right|\right] \tag{6}$$

where n represents the number of samples in the dataset, p represents the number of features, y_i represents is the target value for the ith data point, x_{ij} represents the j-th feature of the i-th sample, β_j represents the coefficient estimates, α is the regularization parameter that controls the strength of the penalty term, and min is the optimization function.

The objective is to minimize this function while obtaining a sparser model by encouraging the coefficients of some features to be exactly equal to zero.

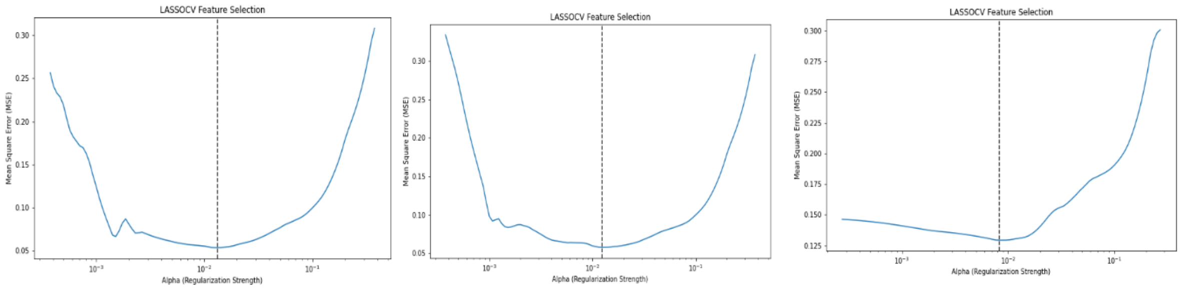
The Lasso-CV algorithm iteratively computes and cross-validates the coefficient vector with varying alpha values, to select a value which gives a good balance between accuracy and model sparseness (i.e. a small number of coefficients other than zero).

Fig. 2 lists the dimensions of the speech features and the features dimensions after Lasso-CV.



	Articulation features	Phonation + Articulation features	DDK features
Total extracted features	80	86	11
Number of seleced features	13	15	9

Fig. 2. The dimension of speech features and their dimensions after using Lasso-CV



(a) (b) (c)

Fig. 3. Estimation of mean square error by regularization strength alpha using a) Articulation features, b) Phonation + Articulation features, c) DDK features

Fig. 3 shows the results of mean squared errors a function of different alpha values to find the optimal alpha for your Lasso regression with a cross-validation model. The optimal choice of selection alpha is 0.013, resulting in the lowest MSE of 0.053 when using articulation features but the optimal selection of alpha is 0.012 when using both phonation features and articulation features, resulting in an MSE of 0.057, the optimal alpha when using DDK features is 0.008 with lowest MSE = 0.129. This specific alpha value produces the most precise and dependable for the feature selection task. This implies that the selected model obtains an optimal balance between maximizing the predictive performance and minimizing the complexity of the feature set.

E. Classification models

Three types of machine learning classification algorithms, Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Machine (SVM) are trained on the dataset to find the best one for the whole model.

The models take the PD dataset including (articulation & articulation+ phonation & DDK) features as input and then in output models extract the status of the patient (PD or HC).

ANN is a computational model inspired by the structure and functioning of the human brain. The network consists of interconnecting artificial neurons that process and analyze information. It has input and output layers, along with one or several hidden layers [3]. Each neuron in the network receives and processes inputs from the previous layer to produce an output. The network enhances its performance over time through a learning process that adjusts the strengths and importance of the connections between the neurons, referred to as weights.

In this study, an Input Layer represents the features and five hidden layers. Then we picked a rectified linear unit (Relu) as an activation function for hidden layers, we adopted a relu to eliminate eigenvalues less than zero at the site to accelerate the model training process between layers. Finally, the activation function of the sigmoid is connected to the classification output.

RF classifier is a machine learning algorithm utilized for classification tasks. It is an ensemble learning technique that combines multiple decision trees to form predictions. Each decision tree within the random forest is constructed using a subset of the training data along with a random selection of features. This randomness helps to mitigate overfitting and increase the classifier's accuracy.

To generate a prediction using a random forest classifier, the input data is fed through each decision tree in the forest, and the final prediction is determined based on the majority vote of the trees. This method enhances the classifier's robustness and accuracy.

SVM classifier is a supervised ML algorithm that categorizes data into different classes. It is a powerful learning method that has been extensively used in various biomedical and health informatics-related problems. The SVM model constructs a hyperplane, or a set of hyperplanes, in a high-dimensional space during the training process to separate the data points into their respective classes [19]. SVM generally aims to maximize the gap between the support vectors, which refer to the training examples nearest to the decision boundary, and the separating hyperplanes. The model consists of hyperparameters such as the choice of kernel, kernel parameters, and regularization parameters. This study used two activation function kernels, radial basis function (RBF) and linear kernel.

F. Performance evaluation

The subsequent step involves evaluating the model's efficiency using a test dataset based on some specific metrics. To evaluate the classification performance in our study, we employed various metrics including accuracy, recall, precision, and the F1_score curve. Choosing the appropriate metrics for assessing the classification models is critical since it impacts how the performance is assessed and compared [20].

Accuracy: it determines the percentage of accurately predicted labels or classes in the dataset. It is derived by dividing the number of accurately predicted instances by the total number of predictions.

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

Precision: It is defined as the ratio of true positive instances to the total number of instances.

$$\text{precision} = \frac{TP}{TP + FP}$$

Recall: Also called sensitivity, this is a measure of how thorough the classifier is defined as the proportion of correctly predicted positive examples to the total number of positive occurrences.

$$\text{recall} = \frac{TP}{TP + FN}$$

F1-score: is the harmonic mean of precision and recall, which provides a balanced measure of the model's performance.

$$\text{F1} - \text{score} = \frac{2.TP}{2.TP + FP + FN}$$

The four possible outcomes are described as follows:

True Negative (TN): The number of instances correctly classified into negative class.

True Positive (TP): The number of instances correctly classified into the positive class.

False Positive (FP): The number of instances that are misclassified as positive.

False Negative (FN): The number of instances misclassified as negative instead of belonging to the target class.

4. Results And Discussion

This paper aims to develop a reliable technique for early detecting Parkinson's disease (PD) through the analysis of vocal samples. This study used the Parkinson's database, consisting of 47 voice signal characteristic listings diagnosed in early stage of the disease, they obtained from 25 PD patients and 22 healthy controls. The different features were examined and their impacts on the models were assessed. The study integrated various machine learning algorithms such as ANN, RF, and SVM. The dataset was divided into two different groups to demonstrate the effectiveness of the proposed method. The first set of samples was used as a training set, while the remaining set was reserved for testing and validation purposes in order to determine the accuracy of the system. In addition, Lasso cross-validation was utilized to identify the optimal features.

Tables I-III display the performance of PD detection using three different classifiers (ANN, RF, SVM) when tested with various features. The most outstanding results are emphasized in bold. For articulation features-based PD detection, the best performance metric values are obtained by using the RF classifier, with an accuracy of 93.33%. When utilizing (articulation + phonation) features, both RF and SVM algorithms achieve the highest accuracy of 93.33%. ANN achieves the highest classification performance, with an accuracy of 98.47%, among all classification models when incorporating DDK features. The superior classifier varies depending on the features used.

Table I. Pd Detection Performance Metrics By Using Articulation Features Without Selected By Lasso-Cv

Model	Accuracy (%)	Precision (%)	Recall (%)	F1_score (%)
ANN	73.68	69.23	90	78.26
RF	**93.33**	87.5	100	93.33
SVM (kernel=RBF)	80	66.67	100	80
SVM (kernel=linear)	73.33	66.67	66.67	67

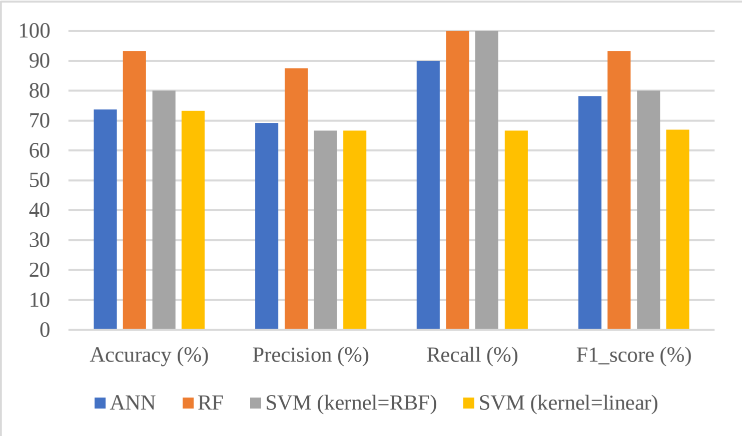


Figure 3. Comparison results of articulation features from different models without using Lasso-CV technique

Table II. Pd Detection Performance Metrics By Using Articulation Features With Phonation Features Without Selected By Lasso-Cv

Model	Accuracy (%)	Precision (%)	Recall (%)	F1_score (%)
ANN	78.95	71.43	100	83.33
RF	**93.33**	100	83.33	90.91
SVM (kernel=RBF)	**93.33**	85.71	100	92.31
SVM (kernel=linear)	80	80	66.67	72.73

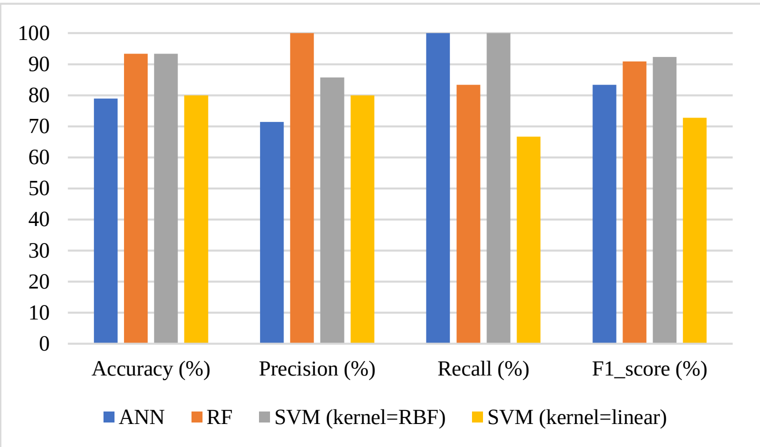


Figure 4. Comparison results of phonation features with articulation features from different models without using Lasso-CV technique

Table III. Pd Detection Performance Metrics By Using Ddk Features Without Selected By Lasso-Cv

Model	Accuracy (%)	Precision (%)	Recall (%)	F1_score (%)
ANN	**89.47**	90	90	90
RF	70	100	57.14	72.73
SVM (kernel=RBF)	73.33	77.78	78	77.78
SVM (kernel=linear)	66.67	75	67	70.6

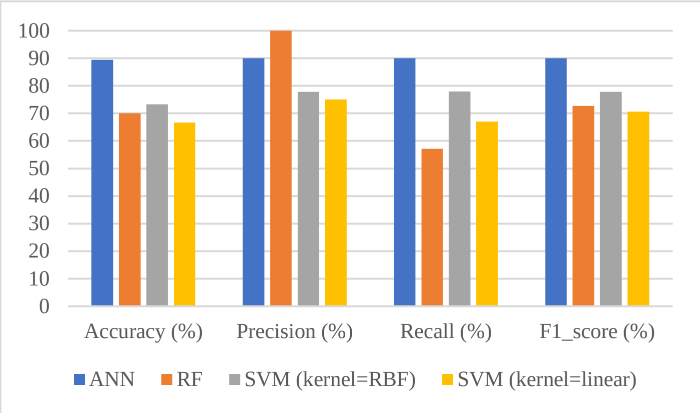


Figure 5. Comparison results of DDK features from different models without using Lasso-CV technique

Tables IV-VI present the results of PD detection using various features obtained through the Lasso-CV feature selection algorithm and classified by three classifiers. It is evident that the SVM and RF models displayed perfect accuracy of 100% when considering articulation only or both articulation and phonation. In contrast, the ANN classification model displayed the lowest accuracy of 94.74%. RBF kernel SVM displayed the highest accuracy of 80%, surpassing all other classifiers when using DDK features as input.

Table IV. Pd Detection Performance Metrics By Using Articulation Features Selected By Lasso-Cv

Model	Accuracy (%)	Precision (%)	Recall (%)	F1_score (%)
ANN	94.74	100	90	94.74
RF	**100**	100	100	100
SVM (kernel=RBF)	**100**	100	100	100
SVM (kernel=linear)	**100**	100	100	100

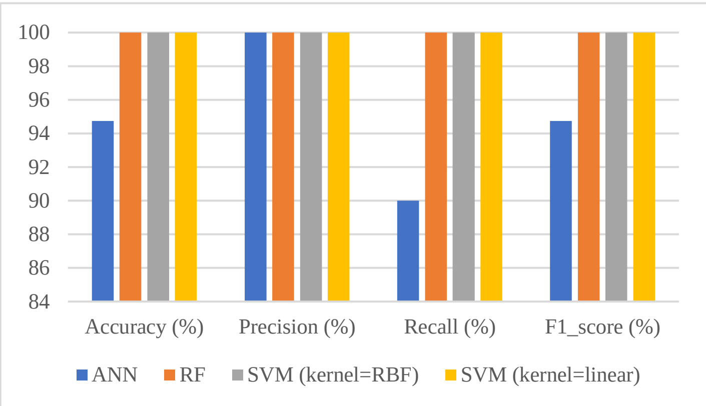


Figure 6. Comparison results of articulation features from different models using Lasso-CV technique

Table V. Pd Detection Performance Metrics By Using Articulation Features With Phonation Features Selected By Lasso-Cv

Model	Accuracy (%)	Precision (%)	Recall (%)	F1_score (%)
ANN	94.74	100	90	94.74
RF	**100**	100	100	100
SVM (kernel=RBF)	**100**	100	100	100
SVM (kernel=linear)	**100**	100	100	100

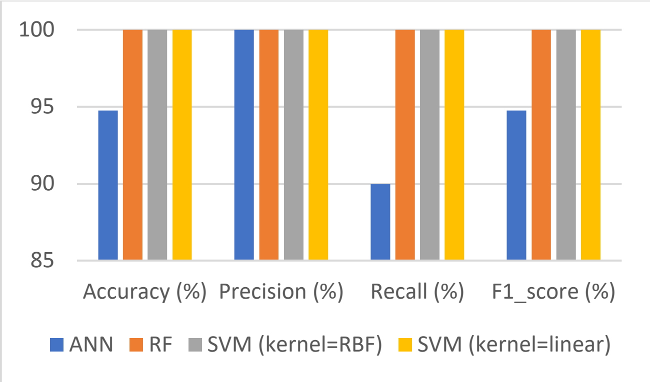


Figure 7. Comparison results of phonation features with articulation features from different models using Lasso-CV technique

Table VI. Pd Detection Performance Metrics By Using Ddk Features Selected By Lasso-Cv

Model	Accuracy (%)	Precision (%)	Recall (%)	F1_score (%)
ANN	63.16	60	63.16	66.67
RF	60	80	57.14	66.67
SVM (kernel=RBF)	**80**	80	88.89	84.21
SVM (kernel=linear)	73.33	85.71	66.67	75

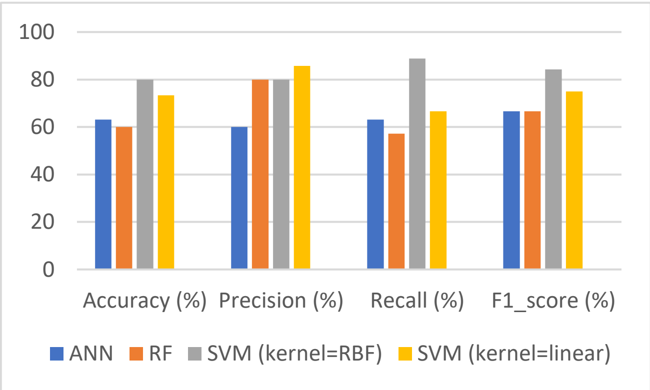


Figure 8. Comparison results of DDK features from different models using Lasso-CV technique

Graphical representations of the accuracy, precision, recall, and F1-score results of algorithms calculated for classifiers (including ANN, RF, and SVM) with and without the selection technique Lasso-CV are presented in Figures 4-9. The classifiers achieved significantly higher accuracy when using articulation features alone or when using both phonation and articulation features, compared to DDK features. This evidence demonstrates the importance of taking articulation features into account when detecting Parkinson's disease, as the classifier achieves superior performance upon their addition.

Comparative Results of Proposed Detection Model with the State-of-the-Art Studies

In this section, we compared the results of this study with those of previous works that made use of the same data and summarized in the table. Table VII presents the comparison of the results of this work with other methods used in previous studies.

Tipathi & Kopparapu [22], train a 1D-CNN model from using IMF signals as raw features. the authors had an accuracy of 84.6%. Moreover, in Toye et al. [23] seven algorithms have been applied to classify the data with Linear Regression and Gradient Boosting feature selection techniques using 24 features. The highest result for that paper exceed 98.9% by SVM model that gives the best result with 10 selected features. To model the loss of motor control and the difficulty to start/stop movements of PD patients, Amato & al. [24] proposed the analysis of the Transition Regions (TR) by extracted 60 features as input to models, they compare classification Accuracy Among 7 classical classifiers by employing a 10-fold CV. The authors had an accuracy of 98% using the SVM classifier. In order to predict Parkinson's disease using voice recordings and deep learning, Maskeliūnas et al. [25] created a deep U-lossian model that is a modified Hybrid Mask U-Net architecture with an adaptive custom loss function. They demonstrated that their technique achieved an accuracy of 89.64%.

Comparison with previous studies demonstrate that our study achieved the highest level of performance and accuracy for detection of Parkinson's disease.

Table VII. Comparison Of The Accuracy Between The Proposed Work And Previous Works

References	Task	Model	Accuracy
Tripathi & al. [22]	Vowel	1D-CNN	84.6%
Toye & al. [23]	Vowel	SVM	98.9%
Amato & al. [24]	Sentences	SVM	98%
Maskeliūnas & al. [25]	Vowel	Hybrid U-Lossian Network	89.64%
Our study	Vowel	RF	100%
	Words	SVM (RBF)	100%
	Syllable	SVM (Linear)	100%

5. Conclusion

In this paper, an effective technique is proposed to discriminate between individuals with early-stage Parkinson's disease and healthy controls by examining their speech samples of vowel phonation, phonetically balanced words, and repetitive syllables. We were able to identify the optimal speech features and best performing machine learning techniques that can be used to predict PD. The best overall classification score achieved 100% accuracy, precision, recall, and F1-score when utilizing either articulation features or (phonation + articulation) features. The findings suggest that articulation features were superior compared to other features. Additionally, the SVM and RF models attained exceptional speed and accuracy performance by using Lasso-CV feature selection technique in comparison to prior research. The study has demonstrated that articulation features are superior in detecting PD-related dysarthria compared to phonation and DDK features among all the classifying algorithms. The proposed research for early detection Parkinson's disease illustrates that voice analysis can provide an inexpensive and non-invasive approach since it only requires audio recordings with basic equipment. Additionally, this model may prove to be an essential diagnostic tool for physicians.

Refrences

[1] A.-M. Tăuţan, B. Ionescu, et E. Santarnecchi, « Artificial intelligence in neurodegenerative diseases: A review of available tools with a focus on machine learning techniques », *Artif. Intell. Med.*, vol. 117, p. 102081, juill. 2021, doi: 10.1016/j.artmed.2021.102081.

[2] Roland, V. (2019). "Production de la parole dans la maladie de Parkinson. Aspects phonétiques, répercussions communicationnelles et perspectives de prise en charge". Unpublished doctoral thesis, Université de Mons. Jury: Harmegnies, B. (Promotor), Cheron, G., Lahaye, W., Hirsch, F., & Abou Haidar, L.

[3] C. Quan, K. Ren, et Z. Luo, « A Deep Learning Based Method for Parkinson’s Disease Detection Using Dynamic Features of Speech », *IEEE Access*, vol. 9, p. 10239-10252, 2021, doi: 10.1109/ACCESS.2021.3051432.

[4] G. Fagherazzi, A. Fischer, M. Ismael, et V. Despotovic, « Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice », *Digit. Biomark.*, vol. 5, nº 1, p. 78-88, avr. 2021, doi: 10.1159/000515346.

[5] J. R. Orozco-Arroyave *et al.*, « NeuroSpeech: An open-source software for Parkinson’s speech analysis », *Digit. Signal Process.*, vol. 77, p. 207-221, juin 2018, doi: 10.1016/j.dsp.2017.07.004.

[6] P. Fan, « Random Forest Algorithm Based on Speech for Early Identification of Parkinson’s Disease », *Comput. Intell. Neurosci.*, vol. 2022, p. 1-6, mai 2022, doi: 10.1155/2022/3287068.

[7] L. Jeancolas *et al.*, « X-Vectors: New Quantitative Biomarkers for Early Parkinson’s Disease Detection From Speech », *Front. Neuroinformatics*, vol. 15, p. 578369, févr. 2021, doi: 10.3389/fninf.2021.578369.

[8] R. Kumari et P. Ramachandran, « A Review on Early Diagnosis of Parkinson’s Disease Using Speech Signal Parameters Based on Machine Learning Technique », in *Futuristic Communication and Network Technologies*, N. Subhashini, Morris. A. G. Ezra, et S.-K. Liaw, Éd., in Lecture Notes in Electrical Engineering. Singapore: Springer Nature, 2023, p. 217-233. doi: 10.1007/978-981-19-8338-2_18.

[9] Q. Wang *et al.*, « Early detection of Parkinson’s disease from multiple signal speech: Based on Mandarin language dataset », *Front. Aging Neurosci.*, vol. 14, p. 1036588, nov. 2022, doi: 10.3389/fnagi.2022.1036588.

[10] S. Dheer, M. Poddar, A. Pandey, et S. Kalaivani, « Parkinson’s Disease Detection Using Acoustic features from Speech recordings », in *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, Bengaluru, India: IEEE, janv. 2023, p. 1-4. doi: 10.1109/IITCEE57236.2023.10090464.

[11] K. M. Alalayah, E. M. Senan, H. F. Atlam, I. A. Ahmed, et H. S. A. Shatnawi, « Automatic and Early Detection of Parkinson’s Disease by Analyzing Acoustic Signals Using Classification Algorithms Based on Recursive Feature Elimination Method », *Diagnostics*, vol. 13, nº 11, Art. nº 11, janv. 2023, doi: 10.3390/diagnostics13111924.

[12] A. Tsanas, M. A. Little, C. Fox and L. O. Ramig, "Objective Automatic Assessment of Rehabilitative Speech Treatment in Parkinson's Disease," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 22, no. 1, pp. 181-190, Jan. 2014, doi: 10.1109/TNSRE.2013.2293575.

[13] « Italian Parkinson’s Voice and Speech | IEEE DataPort » Disponible sur: https://ieee-dataport.org/open-access/italian-parkinsons-voice-and-speech.

[14] G. Dimauro, V. Di Nicola, V. Bevilacqua, D. Caivano, et F. Girardi, « Assessment of Speech Intelligibility in Parkinson’s Disease Using a Speech-To-Text System », *IEEE Access*, vol. 5, p. 22199-22208, 2017, doi: 10.1109/ACCESS.2017.2762475.

[15] Zwicker, E. and Terhardt, E., “Analytical expressions for critical-band rate and critical bandwidth as a function of frequency”, Acoustical Society of America Journal, vol. 68, no. 5, pp. 1523–1525, 1980. doi:10.1121/1.385079.

[16] « Learning from Audio: The Mel Scale, Mel Spectrograms, and Mel Frequency Cepstral Coefficients | by mlearnere | Towards Data Science ». Consulté le: 25 octobre 2023. [En ligne]. Disponible sur: https://towardsdatascience.com/learning-from-audio-the-mel-scale-mel-spectrograms-and-mel-frequency-cepstral-coefficients-f5752b6324a8.

[17] « sklearn.linear_model.LassoCV », scikit-learn. Consulté le: 13 septembre 2023. [En ligne]. Disponible sur: https://scikit-learn/stable/modules/generated/sklearn.linear_model.LassoCV.html

[18] D.-H. Shih, C.-H. Liao, T.-W. Wu, X.-Y. Xu, et M.-H. Shih, « Dysarthria Speech Detection Using Convolutional Neural Networks with Gated Recurrent Unit », *Healthcare*, vol. 10, nº 10, Art. nº 10, oct. 2022, doi: 10.3390/healthcare10101956.

[19] R. Alshammri, G. Alharbi, E. Alharbi, et I. Almubark, « Machine learning approaches to identify Parkinson’s disease using voice signal features », *Front. Artif. Intell.*, vol. 6, p. 1084001, mars 2023, doi: 10.3389/frai.2023.1084001.

[20] I. Nissar, D. Rizvi, S. Masood, et A. Mir, « Voice-Based Detection of Parkinson’s Disease through Ensemble Machine Learning Approach: A Performance Study », *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 5, nº 19, p. 162806, août 2019, doi: 10.4108/eai.13-7-2018.162806.

[21] D. Malhotra et G. K. Sodhi, « Hybrid Deep Learning Model for COVID-19 Prediction Using Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (LSTM) Network », *Int. J. Comput. Theory Eng.*, vol. 15, nº 3, p. 125-129, 2023, doi: 10.7763/IJCTE. 2023.V15.1341.

[22] A. Tripathi et S. K. Kopparapu, « CNN based Parkinson’s Disease Assessment using Empirical Mode Decomposition ».

[23] A. A. Toye et S. Kompalli, « Comparative Study of Speech Analysis Methods to Predict Parkinson's Disease ». arXiv, 14 novembre 2021. Consulté le: 23 octobre 2023. [En ligne]. Disponible sur: http://arxiv.org/abs/2111.10207.

[24] F. Amato, L. Borzi, G. Olmo, C. A. Artusi, G. Imbalzano, et L. Lopiano, « Speech Impairment in Parkinson's Disease: Acoustic Analysis of Unvoiced Consonants in Italian Native Speakers », *IEEE Access*, vol. 9, p. 166370-166381, 2021, doi: 10.1109/ACCESS.2021.3135626.

[25] R. Maskeliūnas, R. Damaševičius, A. Kulikajevas, E. Padervinskis, K. Pribuišis, et V. Uloza, « A Hybrid U-Lossian Deep Learning Network for Screening and Evaluating Parkinson's Disease », *Appl. Sci.*, vol. 12, nº 22, p. 11601, nov. 2022, doi: 10.3390/app122211601.