Research Article GU J Sci 38(3): x-x (2025) DOI: 10.35378/gujs.1602747

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

Neuroparknet: A New Deep Neural Network Model For Classification of Parkinson's Disease

Burak CELIK[1,*] , Ayhan AKBAL[2]

[1]Kocaeli University, Faculty of Engineering, Department of Electronic and Communication Engineering, Kocaeli, Türkiye
[2]Fırat University, Faculty of Engineering, Department of Electrical-Electronic Engineering, Elazığ, Türkiye

Highlights

- This paper focuses on classification process for Parkinson disease.
- In this study, a new classification architecture for speech data classification is proposed.
- Study assesses ML algorithms performance for Parkinson disease classification using speech features.

Research Article

Received: 16 Dec 2024
Accepted: 03 Mar 2025

Keywords

Parkinson's disease
Speech analysis
Machine learning
Neural networks

Abstract

In recent years, the volume and variety of biological data being acquired have increased significantly. Among these data types, the diagnosis of Parkinson's disease holds a critical place in medical research. For this study, speech signals were recorded from patients and healthy controls in a controlled environment at the Neurology Department of Fırat University Hospital. 28 healthy controls, 22 Med Off patients and 30 Med On patients constituted our data set. Participants were asked to read a standardized text in a quiet room using a high-quality H1N Zoom microphone. 19 features were extracted from the obtained sounds. The dataset was categorized into three distinct classes: Healthy Control, Med Off (patients without medication), and Med On (patients medication). To evaluate classification performance, we used a three-layer deeep neural network (DNN) model as well as classical machine learning algorithms in MATLAB. Various classification scenarios have been considered, including many different combinations. For benchmarking, the DNN results were compared with those from commonly used algorithms in the literature: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), and Naive Bayes (NB). Furthermore, the DNN model's performance was assessed using the NeuroParkNet architecture. The comparative analysis revealed that the DNN model generally provided a more accurate and efficient classification process. However, in some specific cases, its performance was partially outperformed by traditional classification algorithms. These findings highlight the DNN's potential while also underscoring areas for optimization in Parkinson's disease classification systems. In addition, the effects of pharmacological treatments were also evaluated in this study.

1. INTRODUCTION

The number of people with Parkinson disease (PD) has increased significantly with the aging population and rising life expectancy [1]. There are approximately 150 thousand people suffering from Parkinson's disease in Turkey. Globally, this number is approximately 5 million people [2]. Approximately 90% of patients with Parkinson's disease have been reported to have a speech disorder called hypokinetic dysarthria[3]. This condition is recognised as one of the common symptoms accompanying Parkinson's disease and may lead to significant impairment in speech ability depending on the progressions of the disease[4-5]. In Parkinson's disease (PD), speech is affected by pathological symptoms such as akinesia and hypokinesia, leading to reduced amplitude and automaticity of speech movement [6,7]. Speech impairment, which is frequently observed in Parkinson's patients, is usually characterised by voice tremors, intonation difficulties and articulation disorders [7]. This is a common symptom in the early stages of the disease and can result in a marked reduction in fluency and clarity of speech associated with loss of motor control. Consequently, speech analysis has emerged as a potential, non-invasive, and cost-efficient tool for the early diagnosis of Parkinson's disease [8]. In recent years, Machine Learning has emerged as a

*Corresponding author, e-mail: burak.celik@kocaeli.edu.tr

computational approach that delivers exceptional results in solving complex problems, often surpassing human performance in certain tasks [9].

MATLAB and Machine Learning methods have been frequently utilised to address classification issues and extract features that contribute significantly to these classifications. In clinical applications, Machine Learning has demonstrated its potential not only in the classification of Parkinson's disease but also in other areas [10], such as the diagnosis of lung cancer [11]. Speech classification usually aims to discriminate between Parkinson's patients (PD) and healthy controls (HC) with high accuracy [12], as we do in this study. For this, it is important to extract features from speech signals. Various features extracted from speech signals, such as shimmer and jitter, have been used in machine learning algorithms to classify individuals as having Parkinson's disease or being healthy [13]. Extracting meaningful features from the recorded speech signals remains a critical challenge. To achieve optimal validation outcomes, it is essential to prioritize the selection of features that provide the robust [14] to the classification process. Feature selection plays a pivotal role in constructing a reliable predictive model by eliminating redundant or irrelevant features. This process not only reduces the dimensionality of the dataset but also enhances the overall performance and efficiency of the model [15]. Other ML-based methodologies have been proposed for the detection of PD, such as Electroencephalography [16] or Magnetic resonance imaging [17] studies. However, voice analysis has shown its value as a reliable method [18] and includes a completely non-invasive approach. In the study in [19], 84.21% accuracy, 93% precision, 89% sensitivity, 89% sensitivity, 89% F1-score and 87% AUC were achieved in the classification of Parkinson's Disease (PD) patients using machine learning (ML) models. After the implementation of the approach used in this study, the performance metrics improved, with accuracy increasing to 85.09%, precision to 92%, sensitivity to 91%, F1-score to 89% and AUC to 90%. These results show that the method used in this study significantly improves the classification performance of PD detection from audio recordings. In another study [20], the proposed classifier achieved an accuracy of 98.3% with feature selection and 94.92% without feature selection. In this study [21] the first experiment in where 11 acoustic features were used, Support Vector Machine (SVM) showed the best performance by achieving 87.2% accuracy, 83.3% specificity, 87% precision, 90% sensitivity and 88.6% F1 score. However, Logistic Regression surpassed SVM in terms of sensitivity, reaching 91.1%. In the second experiment, all acoustic and MFCC features were combined to create a feature vector with 24 features. SVM again showed the highest overall performance with 98.3% accuracy, 98.7% specificity, 98% sensitivity, 98.9% precision and 98.4% F1 score. In the third experiment, the top 10 features were selected from the initial 24 feature set. With this reduced feature set, SVM achieved superior results compared to other models with 98.9% accuracy, 99% specificity, 98.8% sensitivity, 99.2% precision and 99% F1 score. While SVM provided the best performance using 10 selected features, not all models benefited from feature reduction. For example, Logistic Regression and Gradient Boosting provided better accuracy when using the full set of 24 features (11 acoustic + 13 MFCC). In this study [22], speech signal features were utilized as inputs to machine learning algorithms, and the resulting classifiers were integrated to enhance the accuracy of Parkinson's Disease (PD) classification. Experimental findings revealed a diagnostic accuracy of up to 95% achieved through these machine learning models. Furthermore, a feature extraction methodology informed by clinical expertise was introduced for analyzing speech signals of the participants.

2. MATERIALS AND METHODS

The primary objective of this study is to evaluate the accuracy of machine learning algorithms and a custom-designed deep neural network in classifying Parkinson’s disease (PD). To ensure high-quality data collection and minimize noise, all speech recordings were conducted in a controlled environment. Data acquisition was carried out at the Neurology Department of Fırat University Hospital using a high-quality H1N Zoom microphone. The microphone was mounted on a tripod and positioned according to each participant’s sitting posture. According to recommendations in the literature [13], a standard distance of 10 cm between the microphone and the participant was maintained throughout the recordings. For consistency, all speech signals were recorded in WAV format. The dataset consisted of three distinct groups:

- Healthy individuals (control group),
- PD patients medication (Med On),

- PD patients without medication (Med Off).

For classification, 15% of the dataset was reserved for testing purposes, while the remaining portion was used for training and validation. Various acoustic features were extracted from the recorded speech signals to facilitate classification. Features such as jitter, shimmer, entropy, log entropy, skewness, kurtosis, and power bandwidth were analyzed in the MATLAB environment to identify distinguishing patterns and differences among the groups. These features were introduced into MATLAB's Classification Learning Toolbox to develop and evaluate machine learning models. The comparative classification performance of the models was analyzed to assess their combined effectiveness in distinguishing between the three groups. This comprehensive approach allowed for a robust evaluation of the models' ability to classify Parkinson's disease accurately under varying conditions.

2.1. Dataset

Table 1 includes speech recordings from 28 healthy controls, 22 Med Off patients and 30 Med On patients, providing a balanced data set for robust comparisons. This distribution enabled a comprehensive evaluation of the effects of pharmacological treatment on vocal characteristics in Parkinson's disease. Participants were asked to read a standardized text, "Jale'nin Dünyası", ensuring consistency across all recordings. The quiet room in which the speech recordings were taken is schematically represented in Figure 1. The dataset was categorized into three groups: healthy controls, PD patients medication (Med On), and PD patients without medication (Med Off). To maintain integrity in data collection, the same recording protocol was followed for all participants, including healthy individuals. Speech samples were recorded in WAV format, selected for its lossless and uncompressed properties, which preserve the integrity of the data during storage and processing. These recordings were then transferred to a computer for detailed analysis and classification tasks using MATLAB. This approach ensured high-quality speech data suitable for comprehensive speech signal analysis and machine learning applications.

Table 1. *Dataset*

No	Class	Participants
1	Healthy Control	28
2	Med Off	22
3	Med On	30
Total		80

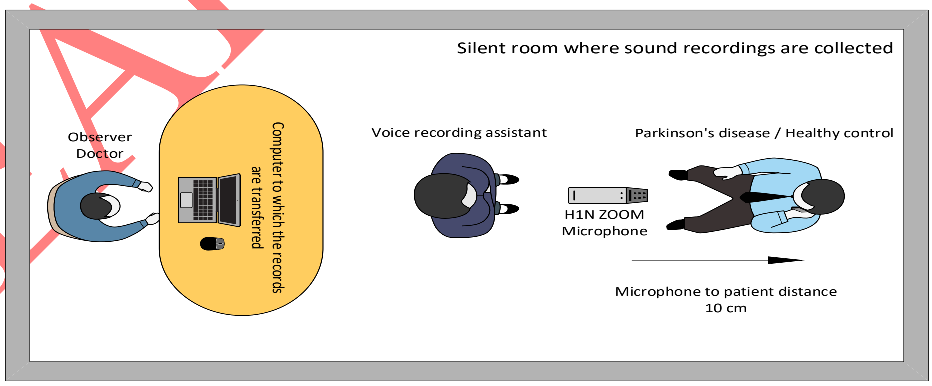

Figure 1. Schematic Representation of the Speech Recording Setup. *This diagram shows the experimental setup used to collect speech recordings from Parkinson's patients and healthy controls. Recordings were performed in a quiet room to minimise external noise and ensure data quality. Participants were seated at a standard distance from a high-quality microphone that captured speech signals while reading a predefined text*

2.2. Statıstıcal Moments Obtained From Speech Signals

To compare speech signals effectively, it is essential to extract certain features from speech. However, not all features are equally effective in distinguishing between signals. Therefore, it is crucial to identify and extract features that improve the classification process and enable a more accurate and significant comparison. In the work of [22], the dataset contained 26 features, including various features such as fundamental frequency, jitter and luminance, as well as additional features such as formants and spectral entropy. In our study, a total of 19 features were extracted from the speech signals in the dataset we created. This feature is shown in Table 2. Chi2 (Chi-square) [23] feature selection algorithm was used due to its high score. In Equation (1), is the equation describing Chi2. Where O_i is the Observation in the classification. E_i are observations in class *i* when there is no relationship between feature and target. Neighborhood Component Analysis (NCA) was used to evaluate the impact of these extracted features on the classification performance. In speech processing, a set of features such as Mel-frequency cepstral coefficients (MFCC) [24], spectral features, formant frequencies or energy-based features are typically extracted. However, not all of these features are equally important for classification or regression tasks. After this selection process, feature selection algorithms such as MRMR, Chi2, ReliefF, and ANOVA were implemented separately in the MATLAB environment, and the resulting feature importance scores were examined. Among these, the Chi2 algorithm gave the highest scores

2.2.1. Chi2 algorithm

Feature scores were analyzed using the Chi2 feature selection method. The contribution of each feature to the classification performance was statistically determined and only the most relevant features were selected, avoiding unnecessary ones. Bayesian Optimization was used to optimize the hyperparameters of the model, speed up the search process and obtain better results using classical machine learning algorithms and the NeuroParknet neural network we built

$$(x)^2 = \sum_{i=0}^{n} \frac{O_i - E_i}{E_i}. \tag{1}$$

The chi-square $(x)^2$ algorithm evaluates the statistical $(x)^2$ measure between each feature and the target variable, selecting the optimal subset of features that achieve the highest $(x)^2$ scores. This process is guided by Equation (1) [25], ensuring that only the most relevant features contributing to the target variable are retained for analysis.

2.2.2. NCA analysis

Neighborhood Component Analysis (NCA) is an advanced metric learning algorithm specifically designed to enhance the classification performance of the stochastic nearest neighbors method. By learning a feature transformation that optimally preserves class separation, NCA improves the accuracy and robustness of nearest neighbors classification [26]. The primary goal of Neighborhood Component Analysis (NCA) is to maximize Leave-One-Out (LOO) classification accuracy by learning a supervised linear transformation within the feature space. Unlike traditional approaches that focus exclusively on predefined similarity metrics, NCA takes a distinctive approach by directly optimizing the feature transformation to enhance LOO performance. This method ensures that the transformed feature space is better suited for class separation. Each data point $X_i \in \mathbb{R}^d$ is linearly transformed as shown in Equation (2), the formula can be seen [25] as given below:

$$Z_i = \mathbf{A}x_i \text{ , } \mathbf{A} \in \mathbb{R}^{d' x d} \tag{2}$$

where **A** is the transformation matrix to be learned.

The probability of point X_i, being assigned to X_j (excluding itself) is defined in Equation (3) as follows, the formula can be seen [25] as given as below:

$$P_{ij} = \begin{cases} \frac{\exp\left(-\|z_i - z_j\|^2\right)}{\sum_{k \neq i} \exp(-\|z_i - z_k\|^2)}, & \text{if } j \neq i \\ 0, & \text{if } j = i. \end{cases} \tag{3}$$

The objective of NCA is to maximize the total probability of correct classification under the LOO scheme, defined as follows in Equation (4), the formula can be seen [25] as given below:

$$\mathrm{L}(\mathbf{A}) = \sum_{\mathrm{i}=1}^{N} \sum_{\substack{J=1 \\ Y_i = Y_j}}^{N} (p_{ij}) \tag{4}$$

where y_i and y_j are the class labels of x_i and x_j, respecticely.

Table 2. The used statistical moments. The table provides clarity by highlighting the statistical features derived from speech signals

No	Feature Moment	No	Feature Moment	No	Feature Moment	No	Feature Moment
1	Maximum	6	Power Band Width	11	Variance	16	Entropy
2	Mean Frequency	7	Jitter	12	Amplitude Mean	17	ZCR
3	Minimum	8	Mean Energy	13	Median	18	Sure Entropy
4	Shimmer	9	Root mean square	14	Skewness	19	Q3-Q1(Interquartile Range)
5	Log Entropy	10	Standard deviation	15	Kurtosis		

To enhance the model's accuracy and reliability, the 10-fold cross-validation method was implemented. This technique involves randomly partitioning the entire dataset into ten subsets, with 10% of the data reserved for testing during each iteration. The process is repeated ten times, and the algorithm's average accuracy is computed across these iterations [26]. By evaluating the model's performance on all subsets, its overall capability across diverse portions of the datsa is assessed. This approach effectively reduces the risks of overfitting or underfitting, ensuring a more robust and generalizable model.

2.2.3. NeuroParkNet architecture

In this study, a deep neural network model was developed to classify the speech recordings of Parkinson's patients into three categories: Healthy, Med off, and Med on. The model architecture was designed to process the extracted features and provide accurate classification results. A total of 19 features were used as inputs to the neural network, including statistical and signal-based parameters such as maximum, minimum, variance, entropy, jitter, shimmer, and mean frequency, which were derived from segmented speech recordings.

As seen in Figure 2, the deep neural network architecture consists of three hidden layers:

- First layer**:** 128 neurons, responsible for initial feature extraction and capturing complex patterns in the data.
- Second layer: 64 neurons, performing further abstraction by refining the learned feature representations.
- Third layer**:** 32 neurons, serving as a final layer for consolidating the extracted features before classification.

The ReLU (Rectified Linear Unit) [27] activation function was used in each hidden layer to introduce nonlinearity and enhance the model's learning capacity. The network was trained over 1000 iterations, allowing sufficient time for convergence and optimization of the classification task. The input layer of the model accepts a vector of size 19, corresponding to the 19 extracted features. The output layer consists of 3 neurons, each representing one of the classification categories: Healthy, Med off, and Med on. The proposed model effectively maps the 19 input features to the three output categories, utilizing its multi-layer structure and ReLU activation to learn both low-level and high-level feature representations. This architecture, trained over 1000 iterations, enables accurate classification of the audio signals into the respective conditions, providing valuable insights into the vocal characteristics associated with Parkinson's disease and the effects of medication.

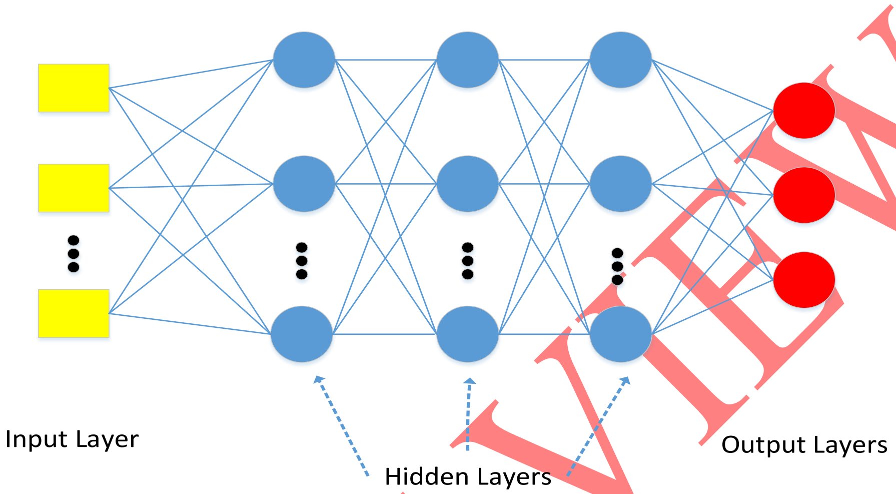

Figure 2*. 3-layer deep neural network architecture*

3. THE RESEARCH FINDINGS AND DISCUSSION

The Classifier Learner in the MATLAB (Matrix Lab.) programming language has allowed many machine learning algorithms to be tested. In the first stage, the features to be obtained from the speech signals were determined and prepared for the analysis process. Neighbourhood Component Analysis (NCA) [18] and Chi2 are used to evaluate the contribution of the extracted features to the overall classification performance. In speech signal processing, a wide range of features are typically extracted, including Mel Frequency Cepstral Coefficients (MFCCs), spectral features, formant frequencies and energy-based features. However, not all of these features have the same level of importance for classification or regression tasks. By identifying and prioritizing the features most effective at distinguishing between classes, NCA serves to eliminate redundant or irrelevant data, thereby optimizing the feature set and reducing its dimensionality. This approach enhances the effectiveness and precision of subsequent modeling processes. To evaluate the differentiation between healthy controls and individuals diagnosed with Parkinson's disease, an initial analysis was conducted focusing on the comparison of Healthy controls, and Medication-OFF state Parkinson's disease cohorts. This investigation aimed to establish a robust framework for employing machine learning-based classification algorithms to achieve accurate predictive outcomes in distinguishing between the two groups. This involves creating a training dataset that allows the algorithm to learn effectively by being provided with relevant data. When extracting features from signals, it is crucial to segment the signal rather than treating it as a whole. Analyzing the entire signal at once leads to the loss of significant variations across the signal's duration, reducing the potential for feature extraction to capture important features. Additionally, the signal lengths obtained from both healthy controls and patients may differ, leading to a dimensionality challenge. Therefore, it is essential to divide the signal into smaller, more manageable segments and extract relevant features from each. This segmentation approach facilitates a more comprehensive representation of the signal's distinct properties, enhancing the classification algorithm's ability to discern patterns and improve prediction accuracy. In this study, signals recorded by a microphone were divided into 3-second segments, followed by the extraction of significant features from each segment for further analysis.

3.1. Classification of Healthy and Med Off

In MATLAB, the features extracted from the signals corresponding to the healthy and med off states, as presented in Table 2, were input into the algorithms in their entirety and subsequently tested. The deep neural network model developed in this study achieved the highest accuracy and showed the highest performance in classifying the given data compared to the other models tested. The confusion matrix for the highest performing model is given in Figure 3.

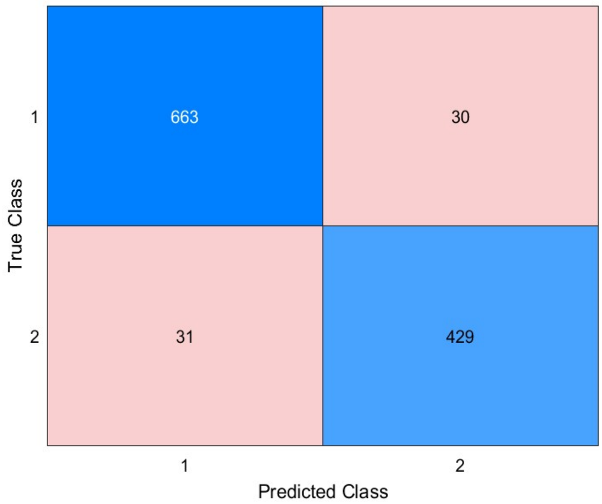

Figure 3. *Confusion matrix for NeuroParkNet Deep Neural Network*

The high rate of correct classification for healthy individuals (Class 1) suggests that the neural network model shows strong overall classification performance. This result indicates the model's high ability to accurately differentiate healthy individuals from other classes. Moreover, the low number of misclassifications further supports the model's capacity to clearly distinguish between classes, minimizing classification errors. Consequently, this reflects an improvement in the model's accuracy, highlighting its enhanced robustness in classification tasks.

Table 3. *Algorithm classification results*

Classification algorithm	MED OFF - 0 %	HEALTHY - 1 %
Decision tree	92.1	87.0
Naive Bayes	87.9	87.4
Support Vector Machine	94.9	92.2
k-NN	92.6	89.8
NeuroParkNet Deep Neural Network	**95.7**	**93.3**

A comparison of the performances between the algorithms is given in Table 3

I. In Decision Tree, 92.1% accuracy was obtained in the Med off class and 87.0% accuracy was obtained in the healthy class,
II. Naive Bayes algorithm, 87.9% accuracy in the Med off class and 87.4% accuracy in the healthy class,
III. Support Vector Machine (SVM) achieved 94.9% accuracy in the Med off class and 92.2% accuracy in the healthy class,
IV. K-Nearest Neighbour (k-NN), 92.6% accuracy in the Med off class and 89.8% accuracy in the healthy class,

V. NeuroParkNet neural network, 95.7% accuracy was obtained in the Med off class and 93.3% accuracy was obtained in the healthy class.

3.2. Classification of Healthy and Med On

After analyzing the Med off condition with healthy controls, the Med on condition was also examined in order to assess the impact of this additional condition. The most effective predictive models for these scenarios were the Support Vector Machine (SVM) and the neural network we developed.

Table 4. *Algorithm classification results*

Classification algorithm	MED ON - 1 %	HEALTHY - 2 %
Decision tree	81.5	84.0
Naive Bayes	82.6	81.1
Support Vector Machine	**90.0**	91.0
k-NN	87.1	88.2
NeuroParkNet Deep Neural Network	89.5	**91.3**

Table 4 presents the accuracy percentages of each algorithm for the Med on and Healthy classes. The decision tree and Naive Bayes algorithms exhibited relatively lower success rates in distinguishing between the Med on and Healthy cases, with accuracies of 81.5% and 82.6%, respectively. In contrast, the Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), and the neural network architecture developed in this study achieved higher accuracy. Notably, the neural network architecture and SVM algorithm demonstrated superior performance, emerging as two of the most successful methods in terms of overall classification accuracy. Figure 4 presents the confusion matrices for two models, respectively.

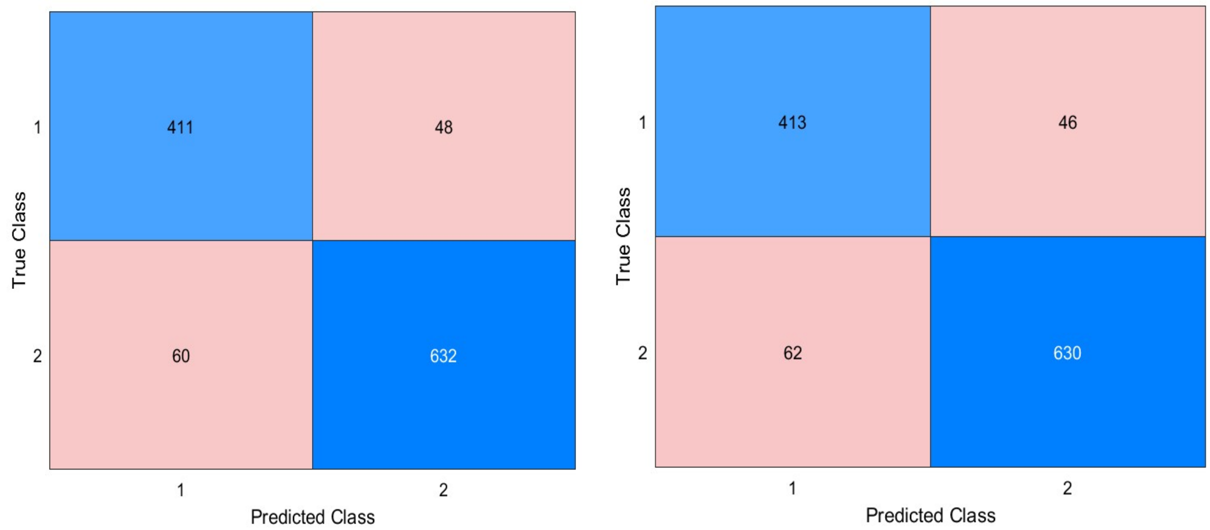

Figure 4. *Confusion matrix for NeuroParknet Deep Neural Network and SVM model respectively (Healthy - Med on)*

3.3. Classification of Med Off and Med On

Classification operations were performed in MATLAB to see the distinction between Med on and Med off. The prediction results of the classification algorithms in Table 5 are evaluated according to the Med off and Med on conditions. The performance of the various machine learning algorithms is given in terms of classification accuracy rates (%). The confusion matrix for the SVM model is given in Figure 5.

Table 5. *Algorithm classification results*

Classification algorithm	MED OFF-1 %	MED ON-2 %
Decision tree	77.3	74.0
Naive Bayes	79.8	58.8
Support Vector Machine	**85.7**	83.2
k-NN	85.0	80.1
NeuroParkNet Deep Neural Network	83.2	**84.8**

I. In Decision Tree, 77.3% accuracy was obtained in the Med off class and 74.0% accuracy was obtained in the healthy class,

II. Naive Bayes algorithm, 79.8% accuracy in the Med off class and 58.8% accuracy in the healthy class,

III. Support Vector Machine (SVM) achieved 85.7% accuracy in the Med off class and 83.2% accuracy in the healthy class,

IV. K-Nearest Neighbour (k-NN) achieved 85.0% accuracy in the Med off class and 80.1% accuracy in the healthy class,

V. NeuroParkNet neural network, 83.2% accuracy was obtained in the Med off class and 84.8% accuracy was obtained in the healthy class.

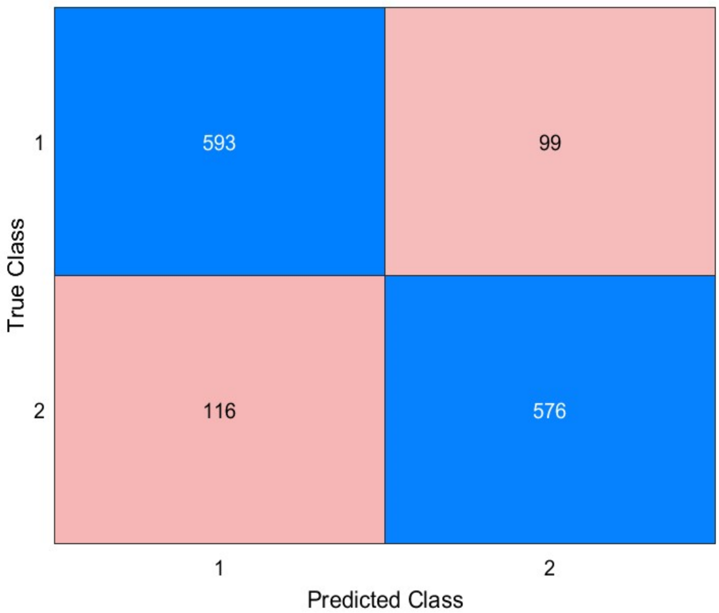

Figure 5. *Confusion matrix for the SVM model (Med off - Med on)*

3.4. Classification of Healthy, Med Off and Med On

The features tested on an individual scale were combined under a table and the algorithms were taught and tested as a whole. The prediction results of the tested audio signals are given in Table 6.

Table 6. *Algorithm classification results*

Classification algorithm	MED OFF- 0 %	HEALTH - 1 %	MED ON- 2 %
Decision tree	74.8	80.5	61.3
Naive Bayes	78.3	82.1	40.4
Support Vector Machine	**82.4**	85.4	75.7
k-NN	81.7	**86.2**	70.9
NeuroParkNet Deep Neural Network	81.2	85.2	**77.3**

The Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), and the neural network architecture demonstrated high performance in identifying Med off sounds, with the SVM achieving the best results among these algorithms. For identifying healthy voices, all algorithms performed well, with the SVM, k-NN, and neural network again showing remarkable accuracy. Notably, the k-NN model exhibited the highest performance in this task. In contrast, for the Med on condition, the neural network model we developed outperformed the other algorithms, highlighting its ability to capture the nuances of this condition. One of the primary reasons for the generally lower performance of the algorithms in the Med on condition may be the increased variability in patients' voices due to the effects of pharmacological treatments. These variations can introduce additional complexity, making it more difficult for models to recognize consistent patterns, thus lowering accuracy. Algorithms like decision tree and Naive Bayes may struggle to adapt to this variability in data distribution, as they might fail to capture the subtle frequency changes or other distinctive features in the patients' signals. Simplified models, such as Naive Bayes, may be particularly limited in capturing these complex relationships between features, further contributing to reduced performance. Consequently, the lower accuracy observed in the Med on cases may be attributed to the individual variability in the effects of the medications on the patients, which poses a challenge for the models in accurately capturing these variations. Figure 6 presents the confusion matrix for the Healthy, Med Off, and Med On states.

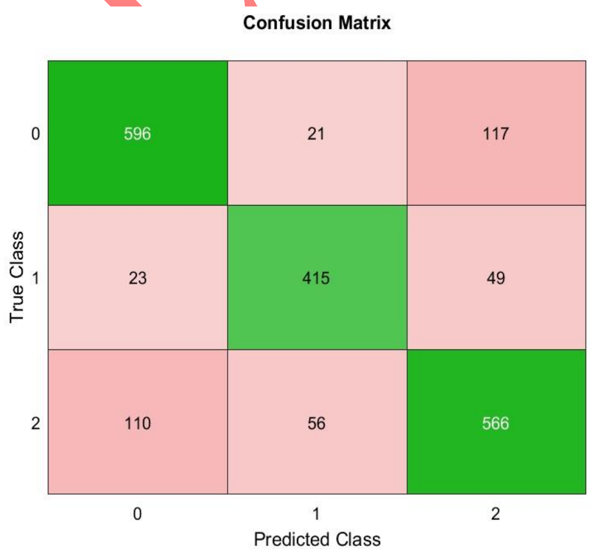

Figure 6. *Confusion matrix for the NeuroParkNet Deep Neural Network (Healthy- Med off - Med on)*

3.5. Healthy and Med (Med off + Med on) Classification

We combined the speech recordings of Parkinson's patients in both the 'Med on' and 'Med off' states into a single dataset. The 'Med on' state refers to recordings taken while patients are on medication, and the 'Med off' state refers to recordings when patients are without medication. By combining these two conditions, we aimed to better understand the effect of medication on the patients' voices.. However, this combination may obscure the distinction between the overall voice profiles of Parkinson's patients and healthy individuals. The features analyzed include parameters extracted from the speech recordings, such as maximum, minimum, variance, entropy, logarithmic entropy, sure entropy, jitter, shimmer, median, power bandwidth, mean frequency, RMS, Q3-Q1 (Interquartile Range), skewness, and kurtosis. These features serve as a benchmark for examining the overall differences in the voices of Parkinson's patients and the impact of medication. Figure 7 presents the confusion matrices for two models, respectively, and Table 7 provides the results of the classification algorithms.

Table 7. *Algorithm classification results*

Classification algorithm	**MED OFF+MED ON - 0 %**	**HEALTHY - 1 %**
Decision tree	92.0	77.0
Naive Bayes	86.1	83.3
Support Vector Machine	**<u>95.0</u>**	**<u>85.4</u>**
k-NN	93.5	79.6
NeuroParkNet Deep Neural Network	93.4	83.7

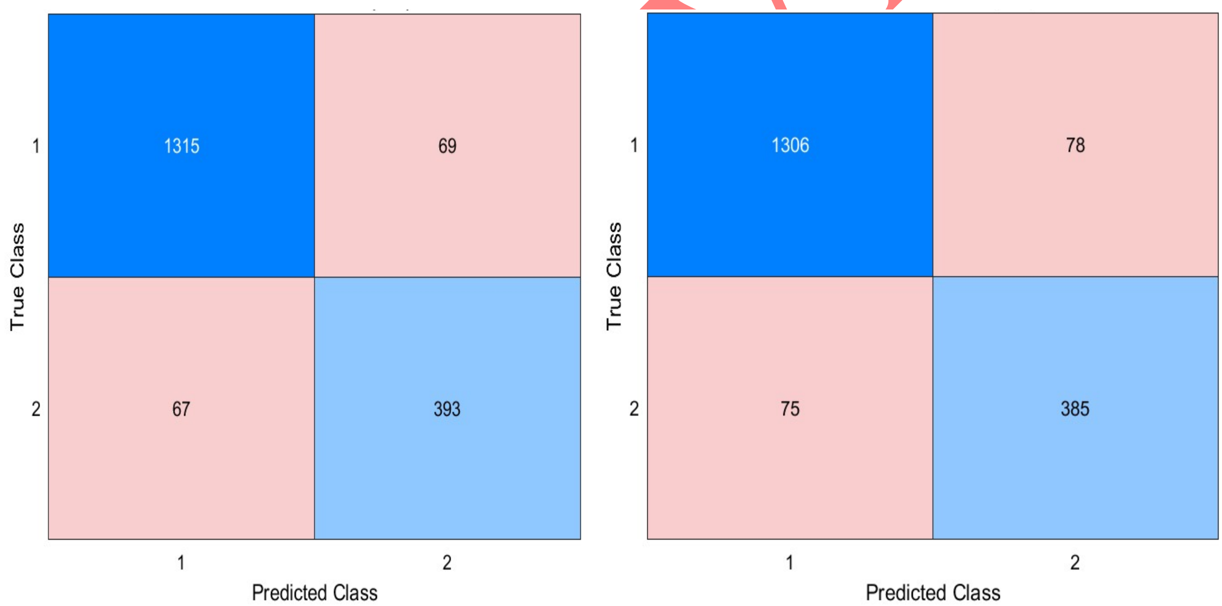

Figure 7. *Confusion matrix for SVM model and NeuroParkNet Deep Neural Network model respectively*

Table 8. *Med (Med Off + Med On) - Healthy confusion matrix parameters (NeuroParknet Deep Neural Network)*

MEASUREMENT	**VALUE**
Recall	94.57
Specificity	83.15
Precision	94.36
Accuracy	**<u>91.70</u>**
F1 Score	94.47

Table 9. *Med Off - Healthy confusion matrix parameters (NeuroParknet Deep Neural Network)*

MEASUREMENT	VALUE
Recall	95.53
Specificity	93.46
Precision	95.67
Accuracy	**94.71**
F1 Score	95.60

As observed in Tables 8 and 9, the combination of Med on and Med off conditions presented difficulties in distinguishing the unique effects of pharmacological treatment. This combination of data made it more difficult to isolate the specific effect of the Med on condition compared to the data presented by the Med off condition. This combination had a significant impact on the predictive performance of the machine learning algorithms. For example, when the Med on and Med off conditions were considered separately, the classification accuracy in the Healthy and Med off scenario was 94.71% (Table 9). However, when the Med on data was integrated with the Med off data, this accuracy dropped to 91.70% (Table 8). Combining these two conditions reduced the distinctiveness of the pharmacological treatment effects and resulted in a slight decrease in the algorithms' ability to distinguish between the conditions. This finding highlights the importance of treating the Med on and Med off conditions as separate conditions when evaluating medication effects in order to preserve the predictive power of the classification models.

4. RESULTS

In this study, first, different conditions were evaluated and then taught to the algorithms on a holistic scale. Then, we aimed to evaluate the effect of medications on the voice characteristics of Parkinson's patients by analyzing the voice recordings in both "Med on" and "Med off" conditions. The dataset was created by combining these two conditions, and thus, a comprehensive analysis of the general profile of Parkinson's patients as well as the effect of medication was examined. Classification performance of various machine learning algorithms, including Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), decision tree, Naive Bayes, and a neural network, was evaluated using the extracted audio features. These features included statistical parameters such as maximum, minimum, variance, entropy, jitter, jitter, median, and others, which provide a basis for distinguishing between Parkinson's patients and healthy individuals.

The study highlights the impact of combining "Med on" and "Med off" conditions on the ability to distinguish healthy individuals from those with Parkinson's disease (PD). The findings are presented as follows:

Healthy vs. Med Off Condition (Table 3):

The NeuroParkNet model achieved the highest classification accuracy, with 95.7% for the Med off condition and 93.3% for the Healthy condition. These results underscore the model's superior ability to differentiate vocal profiles in this context.

Healthy vs. Med On Condition (Table 4):

The NeuroParkNet model again demonstrated robust performance, achieving 91.3% accuracy in classifying Healthy individuals. For the Med on condition, the SVM algorithm performed competitively, reaching an accuracy of 90%. These results illustrate the NeuroParkNet model's capability to maintain high accuracy across different conditions.

Med Conditions Classification (Table 5):

The classification of Med on and Med off conditions was analyzed separately. The SVM model achieved the highest accuracy for the Med off condition (85.7%), while the NeuroParkNet model provided the highest accuracy for the Med on condition (84.8%). This indicates variability in algorithm performance depending on medication states.

Individual Condition Classification (Table 6):

When evaluated independently, SVM outperformed other algorithms in distinguishing the Med off condition, achieving 82.4% accuracy. Conversely, the NeuroParkNet model excelled in identifying the Med on condition, with an accuracy of 77.3%. This demonstrates the nuanced differences in algorithmic strengths for specific conditions.

Combined Med On and Med Off Conditions (Table 7):

When "Med on" and "Med off" data were combined, there was a marked decrease in classification accuracy for distinguishing Healthy individuals. The SVM algorithm outperformed other models, showcasing its ability to generalize across conditions and capture subtle feature differences. The k-NN and decision tree algorithms also performed well but were slightly less accurate compared to SVM and NeuroParkNet. Before combining the conditions, the Healthy-Med off classification achieved an accuracy of 94.71% (Table 9). However, upon combining the conditions, this accuracy dropped to 91.70% (Table 8). This suggests that medication significantly alters the vocal profiles of PD patients, leading to challenges in distinguishing them from healthy individuals. These findings emphasize the importance of accounting for medication effects when developing machine learning models for PD classification. The observed decrease in accuracy with combined datasets indicates that medication may obscure key features in vocal profiles. While SVM demonstrated strong generalization capabilities across conditions, the NeuroParkNet architecture consistently achieved competitive results in most scenarios. Future research should aim to refine models by explicitly addressing the influence of medication on vocal features, potentially through feature selection or advanced normalization techniques.

In conclusion, the study highlights the need to carefully consider medication conditions in PD-related voice analysis to improve classification accuracy and model reliability. Although our NeuroParkNet model blurred accuracy, especially in the Healthy-Med off and Med on+Med off conditions, our NeuroParknet model generally showed high performance in all combinations where Med on was included. These findings underscore the potential of deep neural networks (DNN) in Parkinson’s disease (PD) classification while also highlighting areas for further optimization. Moreover, this study provides an in-depth evaluation of the effects of pharmacological treatments on vocal features, revealing their potential to obscure disease-specific patterns. This emphasizes the importance of addressing medication-induced variability to enhance the reliability and robustness of PD classification systems.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors. The dataset utilized in this research was developed under strict ethical guidelines, ensuring transparency and integrity throughout its application

REFERENCES

[1] Feigin, V. L., Nichols, E., Alam, T., Bannick, M. S., Beghi, E., Blake, N., “Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016”, The Lancet Neurology, 18(5), 459-480, (2019). DOI: https://doi.org/10.1016/s1474-4422(18)30499-x.

[2] Diao, Y., Xie, H., Wang, Y., Zhao, B., Yang, A., Hlavnicka, J., Zhang, J., “Acoustic assessment in mandarin-speaking Parkinson’s disease patients and disease progression monitoring and brain impairment within the speech subsystem”, npj Parkinson's Disease, 10(1), 115, (2024). DOI: https://doi.org/10.1038/s41531-024-00720-3

[3] Muñoz-Vigueras, N., Prados-Román, E., Valenza, M. C., Granados-Santiago, M., Cabrera-Martos,

I., Rodríguez-Torres, J., Torres-Sánchez, I., “Speech and language therapy treatment on hypokinetic dysarthria in Parkinson disease: Systematic review and meta-analysis”, Clinical Rehabilitation, 35(5), 639-655, (2021). DOI: https://doi.org/10.1177/0269215520976267

[4] Arnold, C., Gehrig, J., Gispert, S., Seifried, C., Kell, C. A., “Pathomechanisms and compensatory efforts related to Parkinsonian speech”, NeuroImage: Clinical, 4, 82-97, (2014). DOI: https://doi.org/10.1016/j.nicl.2013.10.016

[5] Müller, J., Wenning, G. K., Verny, M., McKee, A., Chaudhuri, K. R., Jellinger, K., Poewe,W., Litvan, I, “Progression of dysarthria and dysphagia in postmortem-confirmed parkinsonian disorders”, Archives of Neurology, 58(2), 259-264, (2001). DOI: https://doi.org/10.1001/archneur.58.2.259

[6] Rusz, J., Krack, P., Tripoliti, E., “From prodromal stages to clinical trials: The promise of digital speech biomarkers in Parkinson's disease”, Neuroscience & Biobehavioral Reviews, 105922, (2024). DOI: https://doi.org/10.1016/j.neubiorev.2024.105922

[7] Bloem, B. R., Okun, M. S., Klein, C., “Parkinson's disease”, The Lancet, 397(10291), 2284-2303, (2021). DOI: https://doi.org/10.1016/S0140-6736(21)00218-X

[8] Sakar, C. O., Serbes, G., Gunduz, A., Tunc, H. C., Nizam, H., Sakar, B. E., Tütüncü, M., Aydin,T., Isenkul, M. E., Apaydin, H., “A comparative analysis of speech signal processing algorithms for Parkinson’s disease classification and the use of the tunable Q-factor wavelet transform”, Applied Soft Computing, 74, 255-263, (2019). DOI: https://doi.org/10.1016/j.asoc.2018.10.022

[9] Laudis, L. L., Jambek, A. B., A, Lenin Fred., “A Nature Inspired Optimization Algorithm for Parkinson's Disease Classification Through Speech Analysis”, Procedia Computer Science, 235, 840-851, (2024). DOI: https://doi.org/10.1016/j.procs.2024.04.080

[10] Tsanas, A., Little, M., McSharry, P., Ramig, L., “Accurate telemonitoring of Parkinson’s disease progression by non-invasive speech tests”, Nature Precedings, 1-1, (2009). DOI: https://doi.org/10.1038/npre.2009.3920.1

[11] Taye, M. M., “Understanding of machine learning with deep learning: architectures, workflow, applications and future directions”, Computers, 12(5), 91, (2023). DOI: https://doi.org/10.3390/computers12050091

[12] Porumb, M., Stranges, S., Pescapè, A., Pecchia, L., “Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG”, Scientific reports, 10(1), 170, (2020). DOI: https://doi.org/10.1038/s41598-019-56927-5.

[13] Kadir, T., Gleeson, F., “Lung cancer prediction using machine learning and advanced imaging techniques”, Translational lung cancer research, 7(3), 304, (2018). DOI: https://doi.org/10.21037/tlcr.2018.05.15

[14] Saravanan, S., Ramkumar, K., Adalarasu, K., Sivanandam, V., Kumar, S. R., Stalin, S., Amirtharajan, R., “A systematic review of artificial intelligence (AI) based approaches for the diagnosis of Parkinson’s disease”, Archives of computational methods in engineering, 29(6), 3639-3653, (2022). DOI: https://doi.org/10.1007/s11831-022-09710-1

[15] Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgen, F., Delil, S., Kursun, O., “Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings”, IEEE journal of biomedical and health informatics, 17(4), 828-834, (2013). DOI: https://doi.org/10.1109/JBHI.2013.2245674

[16] Escobar-Grisales, D., Ríos-Urrego, C. D., Orozco-Arroyave, J. R., "Deep learning and artificial intelligence applied to model speech and language in Parkinson's disease", Diagnostics, 13(13), 2163, (2023). DOI: https://doi.org/10.3390/diagnostics13132163.

[17] Singh, K. P., Basant, N., Gupta, S., "Support vector machines in water quality management", Analytica Chimica Acta, 703(2), 152-162, (2011). DOI: https://doi.org/10.1016/j.aca.2011.07.027

[18] Loh, H. W., Ooi, C. P., Palmer, E., Barua, P. D., Dogan, S., Tuncer, T., Baygin, M., Acharya, U. R., "GaborPDNet: Gabor transformation and deep neural network for Parkinson's disease detection using EEG signals", Electronics, 10(14), 1740, (2021). DOI: https://doi.org/10.3390/electronics10141740

[19] Kaplan, E., Altunisik, E., Firat, Y. E., Barua, P. D., Dogan, S., Baygin, M., Demir, F. B., Tuncer, T., Palmer, E., Tan, R.S., Yu, P., Soar, J., Fujita, H., Acharya, U. R., "Novel nested patch-based feature extraction model for automated Parkinson's Disease symptom classification using MRI images", Computer Methods and Programs in Biomedicine, 224, 107030, (2022). DOI: https://doi.org/10.1016/j.cmpb.2022.107030

[20] Tuncer, T., Dogan, S., "A novel octopus based Parkinson's disease and gender recognition method using vowels", Applied Acoustics, 155, 75-83, (2019). DOI: https://doi.org/10.1016/j.apacoust.2019.05.019

[21] Hossain, M. A., Amenta, F., "Machine learning-based classification of parkinson's disease patients using speech biomarkers", Journal of Parkinson's Disease, 14(1), 95-109, (2024). DOI: https://doi.org/10.3233/JPD-230002

[22] Kadhim, M. N., Al-Shammary, D., Sufi, F., "A novel voice classification based on Gower distance for Parkinson disease detection", International Journal of Medical Informatics, 191, 105583, (2024). DOI: https://doi.org/10.1016/j.ijmedinf.2024.105583

[23] Toye, A. A., Kompalli, S., "Comparative study of speech analysis methods to predict parkinson's disease", arXiv preprint arXiv:2111.10207, (2021). DOI: https://doi.org/10.48550/arXiv.2111.10207

[24] Yuan, L., Liu, Y., Feng, H. M., "Parkinson disease prediction using machine learning-based features from speech signal", Service Oriented Computing and Applications, 18(1), 101-107, (2024). DOI: https://doi.org/10.1007/s11761-023-00372-w

[25] Yadav, S., Singh, M. K., Pal, S., "Artificial intelligence model for parkinson disease detection using machine learning algorithms", Biomedical Materials & Devices, 1(2), 899-911, (2023). DOI: https://doi.org/10.1007/s44174-023-00068-x

[26] Sidhu, M. S., Latib, N. A. A., Sidhu, K. K., "MFCC in audio signal processing for voice disorder: a review", Multimedia Tools and Applications, 1-21, (2024). DOI: https://doi.org/10.1007/s11042-024-19253-1

[27] Agarap, A. F., "Deep learning using rectified linear units (relu)", arXiv preprint arXiv:1803.08375, (2018). DOI: https://doi.org/10.48550/arXiv.1803.08375