World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Speech-based biomarkers for Parkinson's disease detection and classification using AI Approach

Generaldo Maylem [1,*], Genica Lynne Maylem [1], Isaac Angelo M. Dioses [2], Loida Hermosura [3], James Bryan Tababa [5], Aldrin Bryan Tababa [4], Marc Zenus Labuguen [5] and Dave Miracle Cabanilla [5]

[1] *College of Medicine, Isabela State University, Echague Campus, Echague Isabela, Philippines.*
[2] *School of Information Technology, MAPUA University, Makati Campus, Makati, Philippines.*
[3] *College of Information Technology, Northeastern College, Santiago City, Isabela, Philippines.*
[4] *Philippine Normal University, Alica, Isabela, Philippines.*
[5] *College of Computing Studies, Information and Communication Technology, Isabela State Univeirsty, Cauayan Campus, Cauayan City, Isabela.*

World Journal of Advanced Research and Reviews, 2025, 25(02), 2127-2133

Publication history: Received on 12 January 2025; revised on 23 February 2025; accepted on 26 February 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.25.2.0595

Abstract

Parkinson's disease (PD) is a progressive neurological condition that impairs motor and speech function. Early and precise detection is critical for prompt intervention and illness treatment. This work applies machine learning approaches to classify Parkinson's disease using speech biomarkers collected from voice recordings. The dataset includes a variety of acoustic parameters that capture speech anomalies often seen in people with Parkinson's disease. The Chi-Square (Chi2) approach was used to pick the most important predictors, which improved model performance and reduced computational complexity. The fine K-Nearest Neighbors (KNN) classifier was implemented, achieving a validation accuracy of 74.7%. The model demonstrated a moderate ability to distinguish between Parkinson's and non-Parkinson's cases, as indicated by an area under the curve (AUC) score of 0.7421. However, the confusion matrix revealed challenges in misclassification, with false positives leading to potential unnecessary medical evaluations and false negatives resulting in missed diagnoses. This study highlights the potential of machine learning in Parkinson's detection while emphasizing the need for further refinement to enhance classification accuracy

Keywords: Parkinsons; Machine Learning; Feature Importance; Neurology

1. Introduction

Parkinson's disease (PD) is a progressive neurological disorder that affects millions of people worldwide, with motor and non-motor symptoms including tremors, stiffness, bradykinesia, and voice abnormalities [1]. Early and correct diagnosis is critical for successful illness management because it slows disease progression and improves patient outcomes [2]. Speech-based biomarkers have emerged as a viable technique for Parkinson's disease diagnosis and classification due to its noninvasive nature, low cost, and ease of use. PD patients frequently demonstrate specific vocal deficits such as decreased voice amplitude, monotonicity, and increased jitter and shimmer, which can be recognized utilizing advanced signal processing and machine learning methods [3]. Studies have revealed the ability of speech analysis to diagnose and assess the severity of Parkinson's disease [4]. Also, Bradykinesia, stiffness, resting tremor, and postural and gait abnormalities are examples of motor manifestations of Parkinson's disease (PD); non-motor aspects can include depression, autonomic and olfactory dysfunction, sleep difficulties, mental symptoms, pain, exhaustion, and cognitive impairment [12]. Given the growing recognition of the link between speech impairment and neurodegeneration, speech assessment is becoming an area of study in Parkinson's disease and other degenerative

* Corresponding author: Generaldo Maylem

neurological illnesses. Vocal assessment presents intriguing potential advances because it is inexpensive and non-invasive, and recordings can be made remotely using commonly available smartphone microphones, allowing clinicians to assess speech routinely at the clinic or during patients' daily lives in home and community settings [9,10,11]. Most research on the diagnosis of Parkinson's disease (PD) has relied on sustained phonation of the vowel 'a' as the primary speech task, with some studies also incorporating other vowels such as 'e', 'i', 'o', and 'u'. The features derived from these sustained phonation tasks can be categorized into two types: conventional features and non-conventional features. Conventional features include metrics like jitter, shimmer, harmonics-to-noise ratio, formant bandwidth measurements, and mean formant frequencies. Non-conventional features, on the other hand, involve advanced techniques such as empirical mode decomposition, correlation dimension, and the Hurst exponent. Some studies have utilized a combination of both conventional and non-conventional features to enhance diagnostic accuracy [13,14,15].

Recent advancements in machine learning (ML) have further enhanced the ability to identify subtle speech patterns linked to PD. For example, convolutional neural networks (CNNs) and support vector machines (SVMs) have been successfully employed to classify speech data with high accuracy [5][6]. These models leverage acoustic features such as jitter, shimmer, and harmonic-to-noise ratio (HNR) to distinguish between healthy individuals and PD patients. The integration of large-scale datasets and feature engineering has significantly improved prediction performance, as shown in various studies [7].

However, despite its potential, the adoption of speech-based biomarkers for PD remains challenging due to factors such as data variability, background noise, and limited availability of high-quality datasets [8]. Addressing these challenges requires robust preprocessing, feature extraction, and algorithm optimization techniques.

2. Material and methods

The dataset used in this work is an important resource for studying speech biomarkers related with Parkinson's disease. Voice recordings and a variety of extracted acoustic elements that represent vocal abnormalities frequently seen in people with the condition make up this collection. These characteristics offer insightful information for creating machine learning models intended to detect Parkinson's disease early and track its progression. The dataset is organized to include both patient-specific identifiers and clinically significant speech features. The ground truth for classification tasks is provided by patient data, which includes unique IDs and diagnostic status. Acoustic characteristics include basic frequency measurements, frequency and amplitude changes, and noise-related metrics that reveal vocal instability. To further understand how Parkinson's disease affects vocal function, sophisticated nonlinear dynamical metrics are also incorporated to record abnormalities in speech patterns. The dataset was loaded using Matlab 2024b Software for data analysis process.

2.1. Data processing

The data set used for analysis is secondary and processed. The data set contains no missing or null values. As a result, no data processing occurs to cope with missing and/or null values. To classify PD from this dataset, we divided the data into train and test sets, balanced the dependent value, and chose the features based on the model's specifications. This section discusses data preparation in detail.

To train the Parkinson's dataset, the optimizer settings were tuned with the random search approach. The use of random search enabled effective exploration of the hyperparameter space without relying on established grid arrangements. With the number of iterations set to 30, the model was trained with several randomly chosen hyperparameter combinations, ensuring that the best-performing configuration was determined within the constraints. The acquisition function was set to "expected improvement per second plus," but it remained inactive due to the selected optimizer. This function might have been useful in a Bayesian optimization scenario, but it was not necessary for random search as shown in Figure 1. A training time constraint was not imposed, allowing the optimizer to run all iterations without restriction. If enabled, the training time might have been limited to 300 seconds every run. The number of grid divisions, which was set at 10, was similarly ineffective in this scenario because grid-based procedures were not used. These settings were used to maximize model performance on the dataset, with the goal of improving classification accuracy for Parkinson's disease diagnosis while remaining computationally efficient.

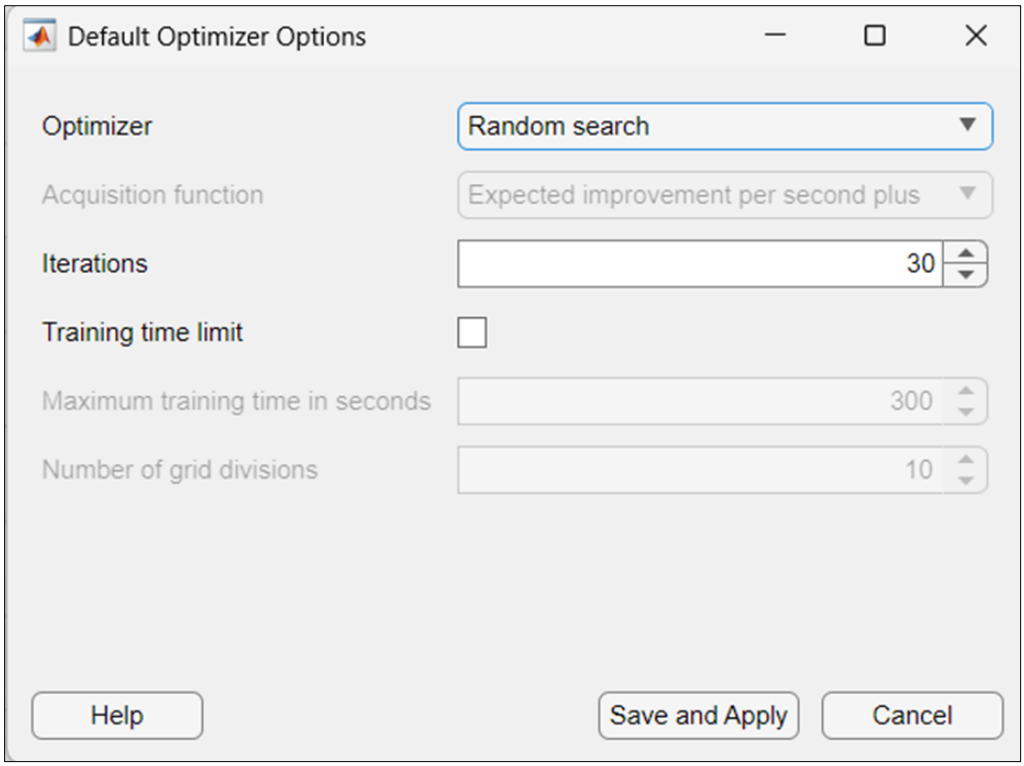

Figure 1 Optimizer Settings

3. Results and discussion

The K-Nearest Neighbors (KNN) classifier was trained on the Parkinson's dataset and achieved a validation accuracy of 74.7%. This accuracy measures the proportion of correctly categorized instances in the validation set, indicating the model's ability to distinguish between Parkinson's and non-Parkinson's cases. The model was configured with fine KNN settings, utilizing all 22 possible features to produce predictions. These traits are expected to include biomedical voice measurements, which are widely employed in Parkinson's disease identification due to their utility in diagnosing motor and speech problems. Fine KNN is a configuration in which the number of neighbors (k) is set to a lower value, which can result in more accurate decision boundaries but also increases sensitivity to noise. The decision to use KNN as a classification method was based on its simplicity and effectiveness in handling small-to-medium-sized datasets without considerable training. However, a validation accuracy of 74.7% indicates that the model's performance could be improved further using further optimization strategies.

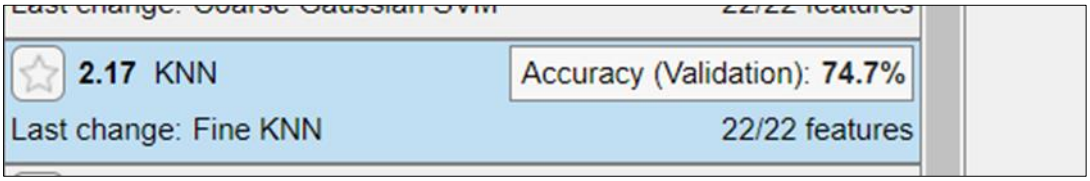

Figure 2 fine knn results

The confusion matrix gives information on the fine K-Nearest Neighbors (KNN) classifier's ability to distinguish between Parkinson's and non-Parkinson's instances. The program properly recognized 339 non-Parkinson's cases and 470 Parkinson's patients, suggesting that it can appropriately categorize most cases. However, it misdiagnosed 161 healthy people as having Parkinson's and failed to discover 113 true Parkinson's cases. The presence of false positives indicates that the model has a tendency to categorize non-Parkinson's cases as Parkinson's, even when the distinction is unclear. This could be owing to overlapping feature values between the two groups, prompting the model to be cautious. While this may be useful in some screening circumstances where early detection is important, it can also lead to undue concern, medical tests, and interventions for those who are healthy. False negatives, though rare, are more significant in a medical setting since they reflect missed Parkinson's diagnosis. Undetected cases may cause delays in therapy and

disease management, which is especially troublesome for a progressive disorder such as Parkinson's. A smaller number of false negatives indicates that the model has an acceptable level of sensitivity, but the fact that some situations are still being missed demonstrates places for improvement. Adjustments to the decision threshold, hyperparameter tweaking, or the incorporation of additional features that better reflect Parkinson's symptoms may improve the model's capacity to distinguish between affected and unaffected individuals. Figure 3 shows the confusion matrix formula and Figure 4 shows the confusion matrix

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{470 + 339}{470 + 339 + 161 + 113} = \frac{809}{1083} \approx 74.7\%$$

Figure 3 Confusion Matrix Formula

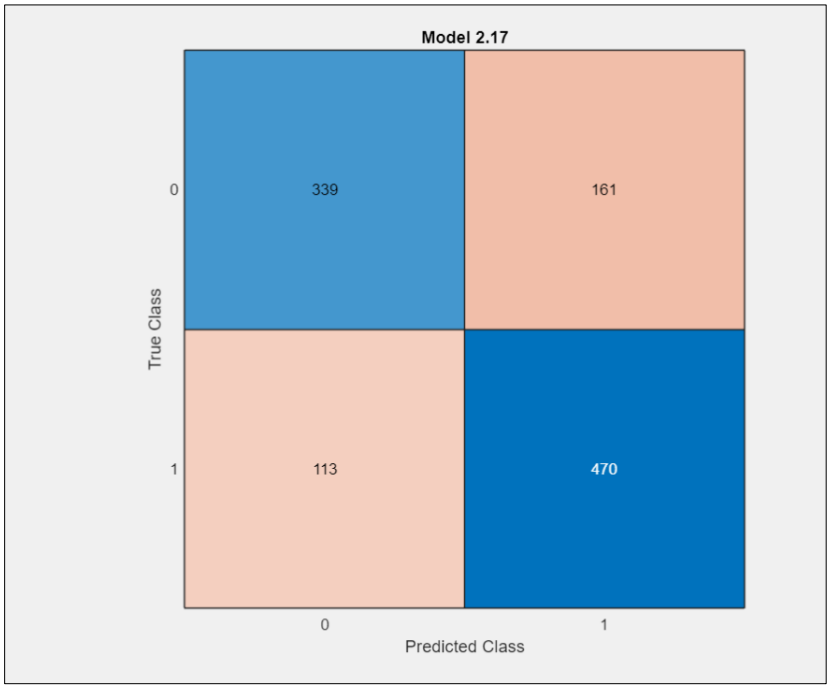

Figure 4 Confusion Matrix

The area under the curve (AUC) is 0.7421, showing that the model has modest capacity to distinguish between the two groups. An AUC of 0.5 indicates that the model performs no better than random guessing, but an AUC closer to 1 suggests great discrimination. With a rating of 0.7421, the model shows good classification skill, but there is potential for improvement. The blue point on the curve denotes the model's operational point, which corresponds to a specific threshold at which the trade-off between true and false positives is established. The strong initial climb in the curve indicates that the model catches a sizable proportion of genuine positives with few false positives at lower thresholds. However, when the false positive rate rises, the true positive rate rises more slowly, indicating diminishing returns. Given the AUC and the shape of the curve, the model is effective but could benefit from refinements such as hyperparameter tuning, feature selection, or alternative distance metrics. Adjusting the decision threshold could also help achieve a better balance between sensitivity and specificity, depending on whether the goal is to minimize missed Parkinson’s cases or reduce false alarms.

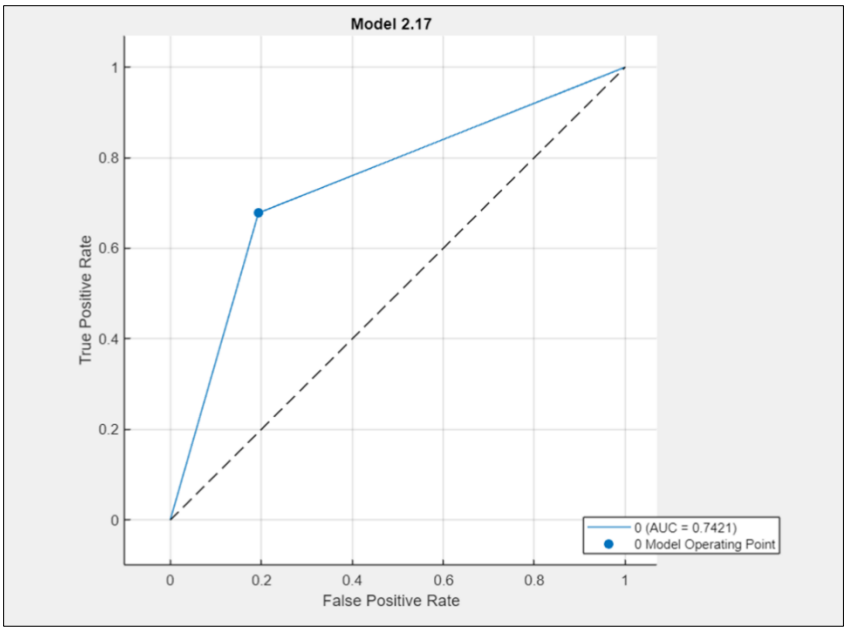

Figure 5 AUC Curve

The Chi-Square (Chi2) method was used to discover the most essential features for classifying Parkinson's cases. The bar chart shows the sorted importance scores of key characteristics, emphasizing their impact on the model's prediction performance. The highest-ranking features, such as MDVPPPQ, MDVPAPQ, and MDVPJitter, show the most significant influence on classification, demonstrating a substantial association with Parkinson's detection.

By selecting features with higher importance scores, the model can potentially improve accuracy while reducing computational complexity. Removing less significant features helps in eliminating noise and preventing overfitting, ensuring that only the most relevant data points contribute to the classification process. This approach aligns with optimizing the fine KNN model, as reducing dimensionality can enhance generalization and improve classification performance. The Chi2 algorithm evaluates feature importance based on statistical dependence, making it suitable for categorical target variables like Parkinson's diagnosis. The selected features primarily relate to vocal impairments, which are key indicators of Parkinson's disease. These include measures of jitter, shimmer, and fundamental frequency variations, reflecting how vocal tremors and instability are captured in the dataset.

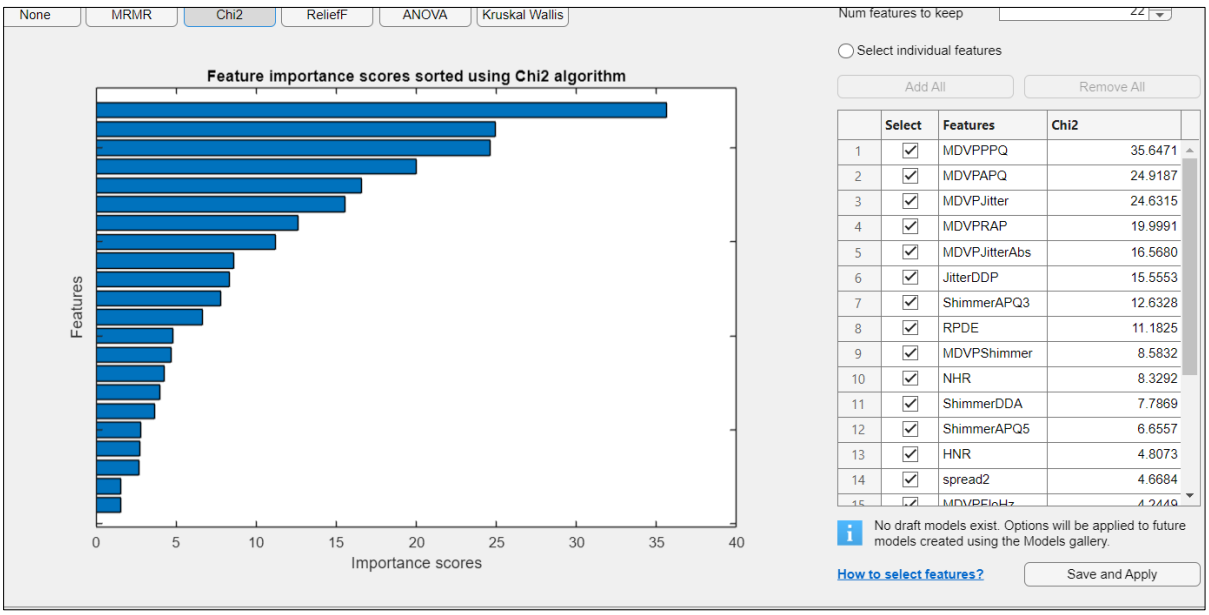

Figure 6 Feature Importance

4. Conclusion

This study utilized machine learning techniques, specifically the fine K-Nearest Neighbors (KNN) classifier, to classify Parkinson's disease based on speech biomarkers. The dataset, rich in acoustic features, provided a valuable foundation for training and testing predictive models. Feature selection using the Chi-Square (Chi2) algorithm helped identify the most relevant vocal attributes associated with Parkinson's, improving model efficiency and interpretability. The fine KNN classifier achieved a validation accuracy of 74.7%, demonstrating its ability to classify Parkinson's and non-Parkinson's cases. However, the confusion matrix highlighted areas for improvement, particularly in reducing false positives and false negatives. The presence of false positives suggests that the model tends to classify non-Parkinson's cases as Parkinson's, which may lead to unnecessary concern. False negatives, while lower in number, are more critical in a medical context as they indicate missed diagnoses. The Receiver Operating Characteristic (ROC) curve and the area under the curve (AUC) score of 0.7421 further confirmed that while the model performed reasonably well, it had limitations in differentiating between the two classes. To enhance classification performance, further optimization techniques such as hyperparameter tuning, alternative distance metrics, and additional feature engineering should be explored. Incorporating more advanced models like Support Vector Machines (SVM) or Neural Networks could also improve predictive accuracy. Additionally, adjusting the decision threshold could help achieve a better balance between sensitivity and specificity, depending on whether the priority is to minimize missed Parkinson's cases or reduce false alarms.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Chen B, Cui W, Wang S, Sun A, Yu H, Liu Y, et al. Functional connectome automatically differentiates multiple system atrophy (parkinsonian type) from idiopathic Parkinson's disease at early stages. Human Brain Mapping. 2023 Jan 20.

[2] Jankovic J. Parkinson's disease: Clinical Features and Diagnosis. Journal of Neurology, Neurosurgery & Psychiatry. 2008 Apr 1;79(4):368–76.

[3] Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. IEEE Transactions on Biomedical Engineering. 2009 Apr;56(4):1015–22.

[4] G. Tsanas et al., "Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease," IEEE Transactions on Biomedical Engineering, vol. 59, no. 5, pp. 1264–1271, 2012.

[5] Eskidere Ö, Ertaş F, Hanilçi C. A comparison of regression methods for remote tracking of Parkinson's disease progression. Expert Systems with Applications. 2012 Apr;39(5):5523–8.

[6] Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, et al. Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings. IEEE Journal of Biomedical and Health Informatics [Internet]. 2013 Jul 1;17(4):828–34. Available from: https://ieeexplore.ieee.org/document/6451090

[7] Nissar I, Mir WA, Izharuddin, Shaikh TA. Machine Learning Approaches for Detection and Diagnosis of Parkinson's Disease - A Review. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). 2021 Mar 19;

[8] Evaldas Vaiciukynas, Adas Gelzinis, Antanas Verikas, Marija Bacauskiene. Parkinson's Disease Detection from Speech Using Convolutional Neural Networks. Springer eBooks. 2018 Jan 1;206–15.

[9] Omberg L, Elias Chaibub Neto, Perumal TM, Pratap A, Tediarjo A, Adams JL, et al. Remote smartphone monitoring of Parkinson's disease and individual response to therapy. Nature Biotechnology. 2021 Aug 9;40(4):480–7.

[10] Zhan A, Mohan S, Tarolli C, Schneider RB, Adams JL, Sharma S, et al. Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity. JAMA Neurology. 2018 Jul 1;75(7):876.

[11] Arora S, Baig F, Lo C, Barber TR, Lawton MA, Zhan A, et al. Smartphone motor testing to distinguish idiopathic REM sleep behavior disorder, controls, and PD. Neurology [Internet]. 2018 Sep 19 [cited 2023 Mar 18];91(16):e1528–38. Available from: https://n.neurology.org/content/91/16/e1528

[12] Kalia LV, Lang AE. Parkinson's disease. The Lancet [Internet]. 2015 Aug;386(9996):896–912. Available from: https://www.sciencedirect.com/science/article/pii/S0140673614613933?via%3Dihub

[13] Peker M, Şen B, Delen D. Computer-Aided Diagnosis of Parkinson's Disease Using Complex-Valued Neural Networks and mRMR Feature Selection Algorithm. Journal of Healthcare Engineering. 2015 Aug;6(3):281–302.

[14] Hariharan M, Polat K, Sindhu R. A new hybrid intelligent system for accurate detection of Parkinson's disease. Computer Methods and Programs in Biomedicine. 2014 Mar;113(3):904–13.

[15] Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease. IEEE Transactions on Biomedical Engineering. 2012 May;59(5):1264–71.