ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14908859

Parkinson's Disease Detection System

Shaba Irram[1] ; Neelam Chakravarti[2] ; Reetesh Gupta[3] ;
S.M. Haider Rizvi[4] ; S. Pratap Singh[5]

[1, 2, 3, 4, 5] Department of Information Technology
Babu Banarasi Das Northern India Institute of Technology

Publication Date: 2025/02/22

Abstract: The Parkinson's disease Detection System is a powerful detection method that can identify whether or not a person has Parkinson's disease. This paper's main objective is to develop a trustworthy system for diagnosing Parkinson's disease that can recognize the condition in a range of patients and enable them to get the help they require quickly. It is thought that if Parkinson's disease symptoms are identified early, a person can receive the necessary medication to keep the condition's negative effects under control. Tremor, sluggish movement, tight muscles, poor posture and balance, loss of instinctive motions, changes in speech or writing, etc. are just a few of the indications and symptoms of Parkinson's disease. Consequently, numerous techniques have been created to test the same for people. As a result, we are developing a detection system that uses several speech metrics to identify Parkinson's disease in a patient.

Keywords: *Parkinson's Disease, Speech Disorders, Dysphonia, Bradykinesia, Cross-Validation, SVM Classifier.*

How to Cite: Shaba Irram ; Neelam Chakravarti ; Reetesh Gupta ; S.M. Haider Rizvi ; S. Pratap Singh (2025) Parkinson's Disease Detection System. *International Journal of Innovative Science and Research Technology*, 10(2), 269-275. https://doi.org/10.5281/zenodo.14908859

I. INTRODUCTION

The condition of nervous system which has impact on the movement of a person can be effected by Parkinson's disease. The first signs that can be noticed at first is the perceptible tremor in any of the hand. After that it shows frequent slow movement which frequently result in tremors. In the early stages the face of the patient may become expressionless and the patient while walk might not swing his arms. His speech could become slurred or dull. In the advance stage, the symptoms of Parkinson's disease get worse. As there is no proper cure for the Parkinson Disease some medicines are helpful in reducing the symptoms if detected on time. In some cases when the situations are critical the doctor may advise for surgery to control the situation and reduce the symptoms.[1,2]

Parkinson's disease symptoms and signs vary from patient to patient. Early symptoms might be slight and go unnoticed. Even after they start to affect both sides of the body, symptoms typically begin on one side and are typically more severe there.[3,4]

Parkinson's disease signs and symptoms include:

- Tremor. A tremor or shaking usually begins in a limb, usually your hand or fingers. A tremor similar to what happens when you rub pills between your thumb and forefinger may occur. Even when your hand is relaxed, it may still tremble.
- Movement slowdown (bradykinesia). Parkinson's disease can cause progressive movement slowdown, which can make even simple tasks challenging and time-consuming. As you walk, your steps might get smaller. Getting out of a chair could be challenging. As you attempt to proceed, you might trip and fall.
- Strong muscles. You can have muscle tightness in any area of your body. Your range of motion may be restricted and made painful by the stiff muscles.
- Poor balance and posture. You can develop stooped posture as a result of it, or you might experience balance issues as a result.
- A slowdown in automatic activity. You could find it harder to walk around with unintentional movements like smiling, blinking, or swinging your arms.
- Disturbances in speech. You can jumble your words, speak quickly, mumble, or hesitate before you speak. Your voice could sound boring because it lacks the usual mannerisms.
- Writing styles differ. Writing could become difficult, and your work can appear crowded.[5,6,7]

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14908859

Parkinson's disease affects more than a million people annually in India. Though this illness cannot be cured, treatment can help. We felt the need to create a machine learning-based Parkinson's disease detection system so that we could diagnose the condition in a range of patients and give them the right care. The main objective of this project is to develop a trustworthy system for diagnosing Parkinson's disease that can detect the condition in a range of patients, enabling them to get the help they require quickly. It is thought that if Parkinson's disease symptoms are identified early, a person can receive the necessary medication to keep the condition's negative effects under control. Jitter, shimmer, NHR, HNR, two nonlinear dynamical complexity measures, average vocal fundamental frequency, maximum vocal fundamental frequency, and minimum vocal fundamental frequency are indicators of Parkinson's disease. Many methods have been developed as a result to test the same for individuals. As a result, we are developing a detection system that uses the numerous indicators stated above to identify Parkinson's disease in a patient.[8][9][10]

II. RELATED WORKS

The aim of this study, according to Arora S et al. [11], was to determine whether voice screening is a feasible population-based method for Parkinson's disease (PD) screening when resources are scarce. They looked at 11,942 sustained vowel /a/ phonations from a US-English set of 1078 PD and 5453 comparison group using the conventional telecommunications network. They used 304 dysphonia metrics to categorise each phonation and measure different vocal deviations. Using 10-fold cross- validation (CV), a balanced sample (n = 3,000 samples) was selected for training and validation, while the other samples (an unbalanced held-out dataset, n = 8942 samples) were used for subsequent model validation. The extreme imbalance of the problem led to the adoption of this approach. Using strict feature selection techniques, they selected 27 dysphonia measurements to feed into a radial- basis-function support vector machine. The results demonstrated 67.43% sensitivity and 67.25% specificity in separating PD participants from controls. These results could open the door to the creation of a low-cost, portable, and reliable diagnostic aid for Parkinson's disease that makes use of speech as a digital biomarker.

Little MA et al. [12] evaluate the usefulness of current conventional and non-standard methods for using dysphonia detection to separate healthy individuals from PD patients. They offer Pitch Period Entropy (PPE), a novel dysphonia measure that is impervious to a number of uncontrollably occurring confounding variables, including loud environments and normal, healthy fluctuations in voice frequency. Thirteen out of the thirty-one participants had Parkinson's disease, and their persistent phonations were recorded. They then selected 10 highly nonstationary measures, and after thoroughly examining every possible aspect of these measures, they selected four that, when merged, yield a kernel support vector machine's approach based classification accuracy of 91.4%. Ultimately, they discovered that non-standard techniques combined with traditional harmonics-to-noise ratios are the most effective means of distinguishing between individuals with Parkinson's disease (PD) and those who are healthy. The non-standard techniques utilised are suitable for telemonitoring applications since they can withstand a variety of unpredictable oscillations in the acoustic environment and individual subjects.

A. Tsanas et al [13] have concentrated on the connection between speech impairments and Parkinson's disease (PD). By using a variety of speech signal processing algorithms, the degree of symptoms associated with Parkinson's disease can now be predicted from speech processing (dysphonia metrics). In this study, they investigate the efficacy of these novel algorithms in distinguishing between PD patients and healthy controls. They use sustained vowels to extract 132 dysphonia measures in total. They then use four feature selection techniques to select four sparse subsets of these dysphonia metrics, and two statistical classifiers—random forests and support vector machines—convert these feature sets to binary classification responses. To demonstrate that these proposed dysphonia metrics perform better than current best practices, they make use of an existing database comprising 263 samples from 43 different people. Using just ten dysphonia traits, they achieve over 99% overall classification accuracy. They observe that a few of the most recently put out dysphonia metrics enhance the ability of classifiers to differentiate between normal participants and PD patients. These results represent a significant advancement in Parkinson's disease noninvasive diagnostic decision support.

In order to determine whether or not the sample belongs to a person with Parkinson's disease, Sakar CO et al. [14] have proposed this work in two ways: (1) to select a minimal set of attributes with the finest combined implications to the PD-score; and (2) to develop a predictive model with the fewest bias (i.e., to maximise the generalisation of the predictions so as to perform well with unseen test examples). The mutual information measure and the permutation test are used to assess the statistical significance and relevance of the relationships between the features and the PD-score for these tasks. The features are then arranged according to the maximum-relevance- minimum-redundancy (mRMR) criterion. A Support Vector Machine (SVM) is used to construct the classification model, which is then validated using a more appropriate cross-validation strategy called leave-one-individual-out, which matches the dataset.

Azadi H. et al. [15] examined the differences between male and female speech characteristics in noisy and quiet settings, as well as the effects of Parkinson's disease on a small set of Jitter and Shimmer voice indicators. 47 samples, consisting of 24 healthy people and 23 patients, were taken from neurology clinics and nursing homes. Separate consideration is given to the best attribute in each category for

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14908859

the samples of men and women. This sentence highlights the vowel "a" as pronounced by the participants. Breakdowns in jitter and shimmer are two of the key elements that are extracted. Using the Relief feature selection method, an ideal pair can be identified in both noisy and quiet environments. According to this study, the Jitter characteristic for Parkinson's disease patients in men and women is 33.4 and 21, respectively. In contrast, the values of the Shimmer feature are 0.06 and 0.1. Moreover, they use only these two features to obtain proper diagnosis rates of 79% and 81% for noisy and noiseless environments, respectively.

A novel and effective technique based on multi-level feature selection was presented by Demir F et al. [16] in order to identify Parkinson's disease from features that included voice recordings of patients who had previously been diagnosed. The Chi-square and L1-Norm SVM algorithms were used to select the first level features (CLS). To increase the sample representation power, the features that these approaches had recovered were combined. The Relief approach was used to select samples from the combined feature set at the top level that had high feature relevance weights. Popular machine learning classifiers, such as KNN, SVM, and DT, were employed in the classification step; the KNN classifier produced the best results. However, by using the Bayesian optimisation algorithm to choose the KNN classifier's hyperparameters, the accuracy of the recommended approach was increased. When the suggested method was tested on a dataset with PD and normal classifications using a 10-fold cross- validation procedure, the classification accuracy was 95.4%.

III. PROPOSED METHODOLOGY

The various steps involved in the making of this project have been mentioned below:

- Downloading Parkinson's dataset
- Doing data pre processing
- Training test data
- Generating new data
- Making use of trained SVM model
- Disease prediction

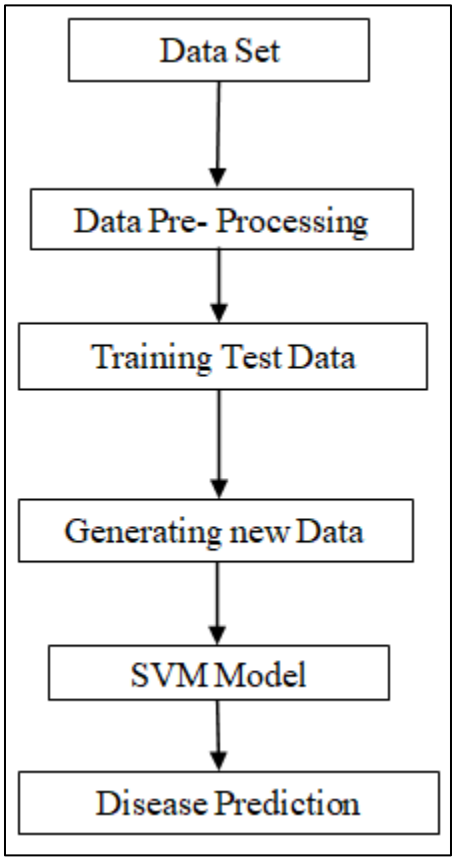

Fig 1: Workflow of Proposed Methodology

A. *Procedure*

- An algorithm integrates software components to recognize numerous symptoms associated with Parkinson's disease in a person.
- The system tracks several biological voice measurements taken from different persons.
- Each column in the table represents a separate voice measure, and each row corresponds to different recordings from different persons.
- The major goal is to differentiate healthy people from those with Parkinson's disease.
- To test the same, the system employs novel machine learning methods.

B. *Parameters Used for Parkinson's Disease Prediction:*

A range of biological voice measurements from 31 patients—23 of whom had Parkinson's disease—were used in the dataset. The table is organised into columns that represent various vocal measures, and rows correspond to one of the 195 voice recordings made by these individuals. The data's main objective is to differentiate between Parkinson's patients and healthy individuals.[17]

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14908859

➢ *Matrix Column Entries (attributes):*

- MDVP:Fo – It is used to measure average vocal fundamental frequency
- MDVP:Fhi – It is used to measure maximum vocal fundamental frequency
- MDVP:Flo - It is used to measure minimum vocal fundamental frequency
- MDVP:Jitter, MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter: DDP – These parameters are used to measure variation in fundamental frequency
- MDVP: Shimmer, MDVP: Shimmer(dB), Shimmer: APQ3, Shimmer: APQ5 MDVP: APQ, Shimmer: DDA – These parametetrs are used to measure variation in amplitude
- NHR, HNR – These two parameters are used to measures of ratio of noise to tonal components in the voice.
- RPDE, D2 – These two parameters are used for nonlinear dynamical complexity measures
- DFA – It is used for Signal fractal scaling exponent
- Spread1, spread2, PPE- These three parametersare used for measures of fundamental frequency variation.

Nowadays, the internet is available all over the world. The hardware required is not expensive. Every member who wishes to obtain information from the website may do so using personal computers, tablets, or other devices. The project is not an expensive product for the customer because the database utilized is MySQL, which is open source. The website can be hosted by one or two technical people. The administrator can upload files (study material) that do not require additional training. The website would feature an intuitive user interface, thus no training would be required to use the product. The patient enrolling to use the website should only be familiar with the English language, basic computer skills, and internet browsing and surfing.

C. Modeling (Analysis & Design)

It includes both project design and a detailed examination of the requirements. In a requirements analysis, the system is examined in light of the requirements of the customer, and the analysis stage establishes the system's starting point, direction of travel, and ending point. System design takes place during analysis-based design.

Streamlit is an open-source Python framework used to develop web applications for data science and machine learning. With Streamlit, creating and launching web apps is simple. You can write Python code in the same way that you would construct an app using Streamlit. The interactive loop of coding and viewing results on the web app is slick thanks to Streamlit.

D. Technology Used

➢ PYTHON - Python is a well-liked, multipurpose, high-level programming language. Python programming is used in web development, machine learning applications, and all cutting edge technologies (Python 3 is the most recent version). Python programming is perfect for both inexperienced and seasoned programmers who use other languages like C++ and Java.

➢ SPYDER - Another well-liked Python Integrated Development Environment (IDE) for developers is called Spyder. It is a Python IDE that is open-source and free, and it has many amazing features like deep analysis, monitoring, and analytical models and debugging. The GUI interface also enables you to locate and edit variables. Spyder comes pre-installed with the Anaconda package manager and is easy to set up, even for novices. Additionally, it works with a variety of scientific Python libraries, such as SymPy, Matplotlib, SciPy, NumPy, and others. That suggests that it was created primarily for engineers and data scientists since it offers a demanding scientific environment for Python development.

➢ SVM (Support Vector Machine): SVM is a component of a new class of learning systems that draws on recent developments in statistical learning theory. The data algorithm is both linear and non-linear. It transforms the initial data into a higher dimension, from which it might use support vectors—essential training tuples— to construct a hyperplane for data separation[18]. When there are more than three features, it is difficult to imagine. For the decision functions, several kernel functions and custom kernels can be provided. A function called the SVM kernel changes a low- dimensional input space into a higher-dimensional space, making a problem that is not separable into one that is. A linear kernel is used when the data can be split into several parts linearly, or along a single line. One of the most often used kernels is this one. When a data set contains a lot of features, it is frequently used.[19]

➢ MACHINE LEARNING - Machine learning is the scientific study of algorithms and statistical models that computer systems use to complete tasks successfully without explicit instructions, instead depending on patterns and inference. It is theoretically classified as artificial intelligence. Through the use of sample data, or "training data," machine learning algorithms create a mathematical model that allows them to make predictions or judgements without explicit guidance. Traditional machine learning algorithms and state-of- the-art deep learning algorithms are the most commonly used techniques for classification

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14908859

tasks. The advantage of the latter is that features can be automatically extracted from the data instead of manually, which greatly avoids errors, however whose implementation requires a huge data set [20].

- STREAMLIT - Streamlit is an open-source app framework with a Python foundation. It makes it simple to develop stunning web apps for data research and machine learning. Many significant Python libraries, including scikit-learn, keras, pytorch, latex, numpy, pandas, and matplotlib, are compatible with it.

- K-fold cross-validation - Cross-validation is a technique for testing machine learning models that involves training multiple models on subsets of the available input data and evaluating them on the complementary subset. Cross-validation can be used to identify patterns that are not able to spread widely or that are overfitting. K groups of comparable-sized samples are created from the input dataset using the K- fold cross-validation technique. Folds are the names given to these samples. The prediction function uses k- 1 folds for each set of training data, and the test set uses the remaining folds. This methodology is popular for resumes because it produces less biassed results than other approaches and is simple to understand.

IV. RESULT ANALYSIS

During this entire procedure, training and testing would be completed exactly once in each set (fold). It aids us in avoiding overfitting. As we all know, the best performance accuracy is obtained when a model is trained using all of the data in a single brief. To overcome this, k-fold cross-validation assists us in developing a generalised model.

To perform this K-Fold Cross Validation, we must divide the data set into three parts: training, testing, and validation, with the volume of data being a challenge.

The Test and Train data sets will aid in the development of models and hyper parameter assessments. In which the model has been validated several times based on the value supplied as a parameter, which is termed K and should be an integer.

To put it simply, the data set would be divided based on the K value, and training/testing would be performed in a time sequence equal to K.

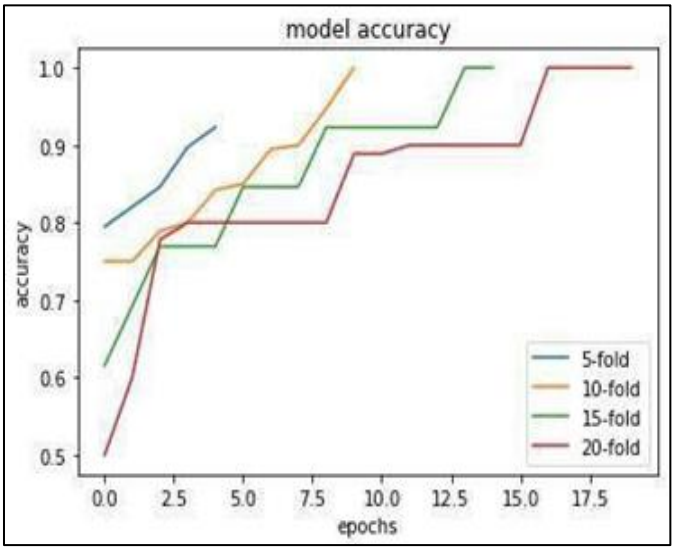

Fig 2 : Accuracy Graph

Here fig.2 represents the accuracy graph between the test data accuracy and number of folds. It displays the findings from the test data that determine how accurately Parkinson disease can be identified for the specified test dataset.

Table 1: Comparision Table

S. No.	Authors	Technique	No. of parameters used	Accuracy obtained
1	Arora S et al.[11]	radial-basis- function support vector machine	27	67.43% sensitivity and 67.25% specificity
2	Little MA. et al [12]	kernel support vector machine	4	91.4%
3	A. Tsanas et al [13]	random forests and support vector machines	4	99.0 %
4	Sakar CO. et al [14]	Support Vector Machine (SVM)	22	92.75%
5	Azadi H. et al [15]	Relief feature selection approach	2	79% for noisy and 81% for noiseless situations
6	Demir F. et al [16]	KNN, SVM, and DT	21	95.4%
7	Proposed Work	Kernel SVM and K-Fold Cross Validation	23	100%

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14908859

Above table 2 is the comparison of techniques and parameters used and accuracy obtained by various researchers for detecting Parkinson disease. The proposed work has achieved 100% accuracy with Kernel SVM with K-Fold Cross Validation Technique. The comparison graph in figure 3 shows the accuracy obtained by various techniques on Parkinson's Disease detection.

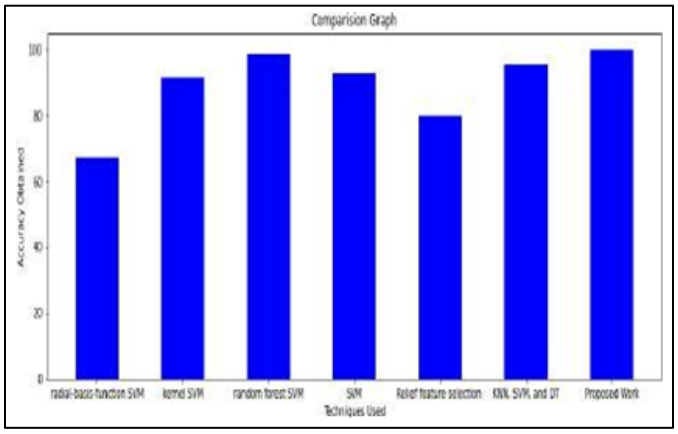

Fig 3 : Comparison Graph

V. CONCLUSION & FUTURE SCOPE

Machine learning techniques have been used to forecast a patient's status and symptoms in order to deliver information that will enhance the patient's condition. It is an example of artificial intelligence in action. Artificial intelligence is a technique that allows computers to learn and develop without being explicitly programmed. Expert systems created using machine learning techniques can help physicians diagnose and predict diseases. Because disease diagnosis is so important to humanity, various studies have been done to establish methods for classification. Although these techniques can be used to predict Parkinson's disease using a set of real-world datasets, the majority of methods developed by supervised prediction techniques in previous studies do not support incremental data updates for PD prediction.

Machine learning techniques have been used to forecast a patient's status and symptoms in order to deliver information that will enhance the patient's condition. Abstract Data mining is the process of extracting usable knowledge from a database in order to create a structure (i.e., model or pattern) that can usefully interpret the data. It has been characterized as the discovery of interesting patterns and information from massive amounts of data. It use machine learning techniques to uncover hidden patterns in data. Expert systems created using machine learning techniques can help physicians diagnose and predict diseases. Because disease diagnosis is so important to humanity, various studies have been done to establish methods for classification.

For future improvement, we will initially try to improve the accuracy of this project and try to make it as accurate as possible. We would also try to provide a testing feature as well so that the users can perform the required test and then use the same data to predict whether they have Parkinson's disease or not.

Declarations of Interest: none

REFERENCES

[1]. Loscalzo J, et al., eds. Parkinson's disease. In: Harrison's Principles of Internal Medicine. 21st ed. McGraw Hill; 2022. https://accessmedicine.mhmedical.com. Accessed April 4, 2022.
[2]. Parkinson's disease: Hope through research. National Institute of Neurological Disorders and Stroke. https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Hope-Through-Research/Parkinsons-Disease-Hope-Through-Research. Accessed April 4, 2022.
[3]. Ferri FF. Parkinson disease. In: Ferri's Clinical Advisor 2022. Elsevier; 2022. https://www.clinicalkey.com. Accessed April 4,2022.
[4]. Chou KL. Diagnosis and differential diagnosis of Parkinson disease. https://www.uptodate.com/contents/search. Accessed April 4, 2022.
[5]. Hornykiewicz O. The discovery of dopamine deficiency in the parkinsonian brain. Journal of Neural Transmission Supplementum. 2006; doi:10.1007/978-3-211-45295-0_3.
[6]. Spindler MA, et al. Initial pharmacologic treatment of Parkinson disease. https://www.uptodate.com/contents/search. Accessed April 4, 2022.
[7]. Relaxation techniques for health. National Center for Complementary and Integrative Health. https://nccih.nih.gov/health/stress/relaxation.htm. Accessed April 4, 2022.
[8]. Taghizadeh M, et al. The effects of omega-3 fatty acids and vitamin E co-supplementation on clinical and metabolic status in patients with Parkinson's disease: A randomized, double-blind, placebo-controlled trial. Neurochemistry International. 2017; doi:10.1016/j.neuint.2017.03.014.
[9]. Tarsy D. Nonpharmacologic management of Parkinson disease. https://www.uptodate.com/contents/search. Accessed April 4, 2022.
[10]. Challa, Kamal & Pagolu, Sasank & Panda, Ganapati & Majhi, Babita. (2016). An Improved Approach for Prediction of Parkinson's Disease using Machine Learning Techniques. International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14908859

[11]. Arora S, Tsanas A. Assessing Parkinson's Disease at Scale Using Telephone-Recorded Speech: Insights from the Parkinson's Voice Initiative. Diagnostics (Basel). 2021 Oct 14;11(10):1892. doi: 10.3390/diagnostics11101892. PMID: 34679590; PMCID: PMC8534584.
[12]. Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng. 2009 Apr;56(4):1015. doi: 10.1109/TBME.2008.2005954. PMID: 21399744; PMCID: PMC3051371.
[13]. A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman and L.O. Ramig, "Novel Speech Signal Processing Algorithms for High- Accuracy Classification of Parkinson's Disease," in IEEE Transactions on Biomedical Engineering, vol. 59, no. 5, pp. 1264- 1271, May 2012, doi: 10.1109/TBME.2012.2183367.
[14]. Sakar CO, Kursun O. Telediagnosis of Parkinson's disease using measurements of dysphonia. J Med Syst. 2010 Aug;34(4):591-9. doi: 10.1007/s10916-009-9272-y. Epub 2009 Mar 14. PMID: 20703913.
[15]. Azadi H, Akbarzadeh-T MR, Shoeibi A, Kobravi HR. Evaluating the Effect of Parkinson's Disease on Jitter and Shimmer Speech Features. Adv Biomed Res. 2021 Dec 25;10:54. doi: 10.4103/abr.abr_254_21. PMID: 35127581; PMCID: PMC8781904.
[16]. Demir F, Siddique K, Alswaitti M, Demir K, Sengur A. A Simple and Effective Approach Based on a Multi-Level Feature Selection for Automated Parkinson's Disease Detection. J Pers Med. 2022 Jan 6;12(1):55. doi: 10.3390/jpm12010055. PMID: 35055370; PMCID: PMC8781034.
[17]. Max A. Little, Patrick E. McSharry, Eric J. Hunter, Lorraine O. Ramig (2008), 'Suitability of dysphonia measurements for telemonitoring of Parkinson's disease', IEEE Transactions on Biomedical Engineering (to appear).
[18]. Bind S., Tiwari A. K., Sahani A. K., Koulibaly P., Nobili F., Pagani M., et al.. (2015). A survey of machine learning based approaches for parkinson disease prediction. Int. J. Comput. Sci. Inf. Technol. 6, 1648–1655.
[19]. Alshammri R, Alharbi G, Alharbi E, Almubark I. Machine learning approaches to identify Parkinson's disease using voice signal features. Front Artif Intell. 2023 Mar 28;6:1084001. doi: 10.3389/frai.2023.1084001. PMID: 37056913; PMCID: PMC10086231.
[20]. M. Chen, Z. Sun, F. Su, Y. Chen, D. Bu and Y. Lyu, "An Auxiliary Diagnostic System for Parkinson’s Disease Based on Wearable Sensors and Genetic Algorithm Optimized Random Forest," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 2254-2263, 2022, doi: 10.1109/TNSRE.2022.3197807.