IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

AN INTERPRETABLE SPEECH ANALYTICS FRAMEWORK FOR EARLY PARKINSON'S DISEASE IDENTIFICATION USING K-NEAREST NEIGHBORS

[1]Ayesha Fathima, [2]Dr. Safia Khanum, [3]Ruqiya Fatima
[1]MTech Scholar, Dept of AI & DS, Shadan Women's College of Engineering and Technology, Hyderabad, ayeshafathima21601@gmail.com
[2]Associate Professor, Dept of CSE, Shadan Women's College of Engineering and Technology, Hyderabad, safiakhanam89@gmail.com
[3]Assistant Professor, Dept of CSE, Shadan Women's College of Engineering and Technology
ruqiyafatima2716@gmail.com

ABSTRACT

Parkinson's disease (PD) is a long-term, degenerative neurological condition that affects the nerve system of the body and causes problems with speech, movement, and coordination. Dysphonia, a voice problem marked by changes in vocal quality, pitch, and loudness, is one of the early signs of Parkinson's disease. Vocal feature analysis becomes a useful method for early identification because between 70 and 90 percent of people with Parkinson's disease have speech problems. Previous research has used a variety of machine learning models to identify Parkinson's disease (PD) using speech signals; nevertheless, issues such class imbalance, optimal feature selection, and limited interpretability are still common. In order to get over these restrictions, this study presents a prediction framework for Parkinson's disease identification based on speech data that makes use of the K-Nearest Neighbours (KNN) algorithm. KNN, a non-parametric and instance-based learning method, classifies patients by comparing their voice feature patterns to those of their nearest neighbours in the dataset. Its simplicity, versatility in dealing with non-linear data distributions, and effectiveness with small to medium-sized datasets make it appropriate for medical diagnostic applications. The algorithm can achieve high accuracy in differentiating PD patients from healthy persons by optimising the value of k and employing distance measures such as the Euclidean or Manhattan distance. Furthermore, feature normalisation and dimensionality reduction techniques are used to enhance KNN's performance and dependability. This strategy attempts to improve the accuracy of early Parkinson's identification while maintaining interpretability, providing a clinically meaningful and data-driven alternative to existing diagnostic procedures.

1. INTRODUCTION

Parkinson's disease (PD) is one of the most prevalent chronic and progressive neurological illnesses, impacting millions of individuals worldwide. It mainly affects the neurological system, gradually weakening motor capabilities like movement, coordination, balance, and speech. Speech impairment is one of the first and most common symptoms, with about 70 to 90 prcent of patients experiencing vocal alterations. These speech-related symptoms, known as dysphonia, include decreased pitch variation, poor vocal quality, and irregular loudness. Detecting such abnormalities early on is critical, as prompt diagnosis and intervention can considerably enhance patient care and quality of life. Traditional diagnostic methods rely mainly on clinical examinations and neurological tests, which may miss early and subtle signs of the disease. As a result, academics are increasingly using data-driven methodologies, notably machine learning, to improve diagnosis accuracy. Machine learning algorithms can analyse vast information, identify hidden patterns, and distinguish between healthy persons and those with Parkinson's disease. Previous research has looked into numerous algorithms for PD detection using speech features; however, issues like class imbalance, overfitting, and interpretability have hindered their practical application. In this situation, K-Nearest Neighbours (KNN) is a potential alternative due to its simplicity and resilience. KNN is a non-parametric, instance-based learning method that classifies new data points according to their proximity to existing samples. It is especially useful for medical diagnostics because of its capacity to manage non-linear interactions and its efficiency with small to medium-sized datasets. Its performance can also be greatly enhanced by optimising parameters like the value of k and the selection of distance metrics. By lowering noise and increasing accuracy, feature normalisation and dimensionality reduction approaches further improve the model's dependability. The suggested KNN framework offers an understandable and effective method for Parkinson's disease early diagnosis by utilising vocal features. Such an approach bridges the gap between conventional diagnostic procedures and contemporary computational techniques, while also providing doctors with data-driven insights. In the end, this study emphasises how crucial it is to integrate machine learning with medical knowledge in order to create scalable, dependable, and accessible Parkinson's disease detection systems.

2. EXISTING SYSTEM OF PROJECT

Conventional machine learning models such as Support Vector Machines (SVM), Decision Trees, Naive Bayes, and Logistic Regression are the mainstay of the current Parkinson's disease diagnosis methods. In order to identify vocal impairments, these models examine speech recordings and extract characteristics including jitter, shimmer, pitch, and harmonics-to-noise ratio. Nevertheless, the majority of these methods struggle to generalise across various datasets and have limited feature selection methodologies. Furthermore, class imbalances—a frequent problem in medical datasets where the number of healthy samples may exceed that of ill individuals—are not well handled by many conventional methods. The lack of interpretability and explainability, which are crucial in clinical applications, is another drawback of current systems. To establish trust and confirm outcomes against clinical expertise, healthcare workers need to understand why a model has produced a certain prediction. The majority of conventional machine learning algorithms have little transparency and do not clearly and practically convey feature value. Because of this, these models frequently fall short in terms of stability, scalability, and clinical applicability, even though they can provide reasonable predictive powers.

2.1. EXISTING SYSTEM DISADVANTAGES

- Inadequate management of class disparity.
- Inadequate generalisation on tiny or noisy datasets
- Naive Bayes is one example of an assumption of feature independence.
- Decision trees, for example, are prone to overfitting.
- Insufficient interpretability of feature relevance
- Exorbitant computational expenses for big datasets
- Ineffective at extracting intricate feature interactions

3. RELATED WORK

Numerous studies have been conducted to forecast patients' Parkinson's disease (PD). The second most prevalent neurodegenerative disease that affects speech and vocal accuracy is Parkinson's disease. Between the ages of 50 and 65 is when Parkinson's disease first manifests. Using a variety of features and methods, numerous research has looked into the use of speech signals for PD diagnosis. This paper looks at recent advancements in using the Explainable balanced Recursive Feature Importance with Logistic Regression (XRFILR) model to address the challenges of speech analysis-based Parkinson's disease diagnosis. It provides creative ways to increase the accuracy and understandability of categorisation.

In the analysis of voice data, machine learning approaches have proven to be quite useful in extracting relevant features for Parkinson's disease monitoring and prediction. Numerous studies have been conducted to forecast patients' Parkinson's disease (PD). The second most prevalent neurodegenerative disease that affects speech and vocal accuracy is Parkinson's disease. Between the ages of 50 and 65 is when Parkinson's disease first manifests. Using a variety of features and methods, numerous research has looked into the use of speech signals for PD diagnosis. In order to address the challenges of diagnosing Parkinson's illness, this study looks at recent advancements in using the Explainable balanced Recursive Feature Importance with Logistic Regression (XRFILR) model to improve categorisation precision and comprehensibility. In order to predict and track Parkinson's disease through speech analysis, machine learning approaches have proven to be very helpful in extracting relevant features from speech data. It provides fresh viewpoints. A deep-learning model (PD) with premotor indicators, such as REM abnormalities, olfactory impairment, cerebrospinal fluid biomarkers, and dopaminergic imaging markers, was introduced by Wang et al. in 2020. With an average accuracy of 96.45% on a dataset comprising 401 early-stage Parkinson's disease patients and 183 healthy individuals, the model outperforms twelve other machine learning methods. Additionally, it improves the interpretability of the model by providing insights about feature significance generated from the Boosting technique. Verifying the model on larger datasets, incorporating additional data kinds, and looking into Parkinson's disease early diagnosis are all lacking.

A deep-learning model for the early detection of Parkinson's disease (PD) was introduced by Wang et al. in 2020. It included premotor indicators such as abnormalities in rapid eye movement (REM), olfactory impairment, cerebrospinal fluid biomarkers, and dopaminergic imaging markers. With an average accuracy of 96.45% on a dataset comprising 401 early-stage Parkinson's disease patients and 183 healthy individuals, the model outperforms twelve other machine learning methods. Additionally, it improves the interpretability of the model by providing insights about feature significance generated from the Boosting technique. Verifying the model on larger datasets, incorporating additional data kinds, and looking into real-time monitoring for continuous early detection are all lacking. In order to identify Parkinson's disease (PD) patients based on their vocal characteristics, Hoq et al. (2021) combined two models and derived from a Support Vector

Machine (SVM) integration utilising Principal Component Analysis (PCA) with a Sparse Autoencoder (SAE). The SMOTE approach was used to equilibrate the PD speech dataset, with an accuracy of 94.4%. PCA and SAE are used in the study to determine whether the effectiveness of identifying the key characteristics for Parkinson's disease diagnosis is inadequate. According to Soumaya et al. (2021), it is important to develop an accurate Parkinson's disease prognostication system. For the Support Vector Machine classifier, the authors used an improved methodology and evaluated the optimisation process of evolutionary algorithms in addition to speech acoustic and decomposition properties. A thorough ensemble methodology was developed by Liu et al. in 2021. Local discriminant Preservation Projection (LPP) places a strong emphasis on PD prediction. While simultaneously increasing the inter-class variance and decreasing the intra-class variance of the samples using an ideal method, the ensemble technique preserves the neighbourhood arrangement of PD voice samples.

PD samples that were out of balance were handled appropriately. In order to counteract the curse of dimensionality, which can also decrease dimensionality, feature selection should have been used. To demonstrate the effectiveness of each classifier, Goyal et al. (2021) assessed a variety of classification techniques. The Significance of Genetic Algorithm (GA) approaches, maximum Relevance Minimum Redundancy (mRMR), and Principal Component Analysis (PCA) were also investigated in order to decrease the cardinality of the voice dataset without significantly compromising precision. Two types of datasets were used. The XG Boost classifier produced increased precision regardless of class imbalance. The possible application of feature selection approaches in the use of vocal features in Parkinson's disease diagnostic procedures is investigated.

Dhar 2022 enhanced the categorization of advanced Parkinson’s disease using a distinctive feature set (40PD, 40 Healthy Control) and the MI-AE-GOLGBM methodology, achieved a performance rate of 84.17%. Future initiatives concentrated on incorporating methods to enhance classification efficacy. Liu et al. 2022 conducted a study on the remote monitoring of patients, employing a feature set comprising 50 attributes. This study employed Shapley Additive explanations with a Light Gradient Boosting Machine, SHAP-LightGBM, demonstrating a novel feature selection and modeling approach. The proposed model achieved a performance outcome of 91.62%. Traditional feature selection technique used. Lamba et al. 2022 evaluated the effectiveness of Naive Bayes, k-NN, and Random Forest (RF) algorithms in diagnosing Parkinson’s disease. The proposed method incorporated a Hybrid Mutual Information Ranking and Feature Elimination MIRFE-XGBoost, show casing a unified feature extraction and selection strategy. The performance outcomes ranged from 95.58% to 95.47%, but lack of transparency in predicting the PD. Hawi et al. 2022 evaluated the efficacy of intense voice therapy in improving functional communication. voice therapy in improving functional communication. Numerous researchers have made significant efforts to diagnose Parkinson’s disease. Alalayah et al. 2023 introduced an innovative approach to enhancing the tactics for early detection of Parkinson’s Disease through speech analysis. Biases may arise if features are unintentionally omitted, excluded, or disproportionately represented. The model proficiently identified the symptoms of PD, exhibiting elevated classification accuracy.

Iyer et al. (2023) analysed spectrograms obtained from voice recordings of people with Parkinson's disease using a pre-trained CNN (Inception V3) with transfer learning. With an AUC of 0.97 for coloured spectrograms and 0.96 for monochromatic spectrograms, the deep learning model improved performance in detecting Parkinson's disease patients. Rahman et al., 2023 use a three-layer Deep Neural Network2 (DNN2), which outperforms all other approaches with an accuracy of 95.41%, to diagnose Parkinson's disease in its early stages. According to this study, DNNs are more effective than traditional machine learning methods when it comes to patient classification using PD obtained from voice data. In order to develop a soft diagnostic tool for detecting Parkinson's disease using vocal signal features, Alshammri et al., 2023 evaluated numerous machine learning models. According to the study, the Support Vector Machines (SVM) model outperformed earlier machine learning models in the diagnosis of Parkinson's disease (PD), attaining an astounding 96% accuracy. Saxena and Andrew 2023 created a classification system for speech input-based Parkinson's disease detection utilising machine learning models like SVM, KNN, Random Forest, and Logistic Regression. It achieves the greatest accuracy of 89.47% on the testing set, demonstrating the effectiveness of Random Forest. Nevertheless, the suggested model ignores explainable insights and class imbalance, which are essential for enhancing the interpretability and credibility of medical diagnosis.

In the framework of nature-inspired techniques, Chawla et al. 2024 presented a technique that makes use of the Zebra Optimisation Algorithm (ZOA) and Recursive Feature Elimination Cross-Validation (RFECV). This approach uses the Novel Information Filter System (NIFS) feature selection to determine PD. Depending on the many characteristics of the dataset, it offers the possibility of identifying other illnesses. Singh and Tripathi 2024 improved the accuracy of diagnosis in the early detection of Parkinson's disease by using voting procedures to determine the final prognosis. Due to the difficulty of training several models at once, this method's dependence on computer resources is a significant drawback. Furthermore, if explainable AI is not used, the accuracy of the model can still depend on the size and quality of the training dataset.

Using SMOTE for data balancing and PCA for dimensionality reduction, Bukhari and Ogudo (2024) presented an ensemble machine learning model that employs the adaBoost classifier to identify Parkinson's disease (PD) from voice signals. It performs exceptionally well as a non-invasive, affordable diagnostic tool (96% accuracy, 98% precision, and AUC 0.99). Except for the limits of the explainable AI component of this work, larger clinical usage is challenged by the model's computational complexity, the need for hyperparameter tweaking, and reliance on a small dataset. The Interpretable Feature Ranking XGBoost (IFRX) model was introduced by Shyamala and Navamani in 2024. It consists of a series of procedures like SVMSmote for class balance, RFE with XGBoost for feature selection with classifier, and Shap for explainability. improved performance, although the authors' study only used a limited dataset. Higher classification scores (85.09% accuracy, 92% precision) were obtained by Hossain and Amenta 2024 using pipelines. Therefore, our study emphasises the potential of machine learning classifiers and pipelines to use speech biomarkers to differentiate Parkinson's disease patients from healthy controls. By effectively addressing the importance of features in high-dimensional data, the pipelining technique increases the accuracy of categorisation. Because the study does not address class imbalance or apply explainable artificial intelligence approaches, its interpretability and treatment efficacy are limited.

According to the literature review above, a number of research have not sufficiently addressed the issue of class imbalance, which may lead to biased model performance that favours the majority class and compromises the validity of the diagnosis. Because there are many related and irrelevant feature subsets, an increase in features results in a drop in model accuracy. In order to include the most relevant attributes, the dataset needs to be improved. According to current research, feature selection strategies are essential for determining the most pertinent features and improving the precision of different machine learning classification algorithms. The benefits of employing explainability techniques, including Shapley Additive Explanations (SHAP), to clarify model predictions and provide insight into the function of specific vocal characteristics in PD classification are highlighted by a number of recent studies. Additionally, group learning Additionally, ensemble learning and hybrid approaches have been effective in reducing class imbalance, a common problem in medical datasets, and improving feature selection to avoid overfitting and duplication.

Here, we suggest an Explainable Recursive Feature Importance with Logistic Regression (XRFILR) model that prioritises choosing the most informative features while equilibrating class distribution. In order to create a more advanced and reliable diagnostic tool, the XRFILR model aims to increase interpretability and accuracy in PD prediction based on voice data. This method uses balanced datasets, recursive feature selection with logistic regression, and interpretability to improve the limitations of previous research and advance the field of speech-based PD diagnosis.

4. PROPOSED SYSTEM

- The related concept of honey encryption, on the other hand, reshapes the distribution of a plaintext so that the cypher text data and the genuine plaintext are identical through a preprocessing step in symmetric encryption
- For instance, when employing AES, the sender and the recipient share the same key in a symmetric encryption system.
- There are two types of deniable encryption: shared key schemes and public key schemes. Let m be the original message that has to be encrypted.
- The one-time pad is a straightforward illustration of deniable encryption. In the sphere of symmetric encryption, ambiguous multi-symmetric cryptography has been established, which is similar to our goals. They cover a variety of attack scenarios, including known-plaintext and selected cypher text.

5. MODULES

1. Accumulating Resources

The collection of the necessary speech datasets from Parkinson's sufferers and healthy people is the main goal of this module. In order to train the model, it entails gathering pertinent voice recordings, feature sets, and metadata. Reliable sources guarantee that a variety of speech features are represented in the dataset. This stage serves as the basis for developing a strong diagnostic framework.

2. Examining Insights:

Here, significant insights regarding voice abnormalities are extracted by analysis of the raw speech data. It involves recognising important characteristics that are frequently impacted in PD patients, such as jitter, shimmer, and pitch fluctuations. The procedure aids in comprehending the connection between disease presence and voice parameters. This guarantees that only the most pertinent characteristics are taken into account for additional investigation.

3. Normalizing Data:

By bringing all features to a consistent scale, normalisation helps to avoid bias in the model. Standardising vocal parameters prevents the dominance of some characteristics because they vary greatly in range. Additionally, this phase enhances the KNN algorithm's efficiency and accuracy. Normalisation improves the fairness of comparisons during classification by making the dataset consistent.

4. Activating the Model:

At this point, the data is classified using the K-Nearest Neighbours (KNN) algorithm. The model determines whether a voice sample belongs to a person with Parkinson's disease (PD) or a healthy person using distance-based comparisons. The model's activation guarantees that the system is prepared for testing and training. This module is essential for turning data into forecasts that can be put into practice.

5. Tuning Hyperparameters:

The main KNN parameters, including the value of k and the selection of distance metrics (Euclidean or Manhattan), are optimised in this module. The accuracy and performance of the model are greatly improved with proper tuning. To determine the best setup, experiments with various configurations are carried out. This guarantees consistent and dependable outcomes from the model.

6. Performance Evaluation:

This module evaluates the model's accuracy and dependability once it has been trained. Effectiveness is measured using metrics including precision, recall, F1-score, and confusion matrix. Evaluation of performance also aids in pinpointing areas that want improvement. This stage confirms the prediction framework's clinical applicability.

7. Calculating Probabilities:

Based on speech characteristics, this last module calculates a patient's probability of having Parkinson's disease. The model gives predictions likelihood scores rather than merely a binary classification. This improves interpretability and helps medical professionals make wise choices. By calculating probabilities, the system's practical usefulness is increased by ensuring that outcomes are delivered with confidence levels.

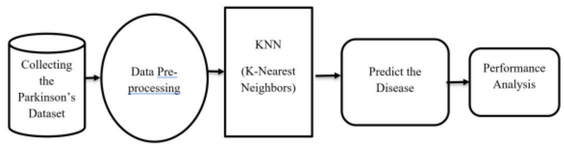
6. TECHNIQUES USED IN PROJECT

6.1. EXISTING TECHNIQUE

Parkinson's disease identification utilising voice features has made substantial use of traditional machine learning methods such as Support Vector Machine (SVM), Decision Tree, Naive Bayes, and Logistic Regression. To distinguish between healthy and afflicted people, these models use extracted auditory characteristics as jitter, shimmer, pitch, and harmonic-to-noise ratio. For instance, SVM is suitable for the PD vs. non-PD case since it performs well in binary classification tasks and high-dimensional domains. Conversely, logistic regression is important for medical diagnostics since it provides interpretable data and is utilised for probabilistic predictions. Nevertheless, class imbalance, overfitting on short datasets, and inconsistent feature value evaluation are some of the problems that these current models frequently face. In correlated vocal data, Naive Bayes' assumption of feature independence might not hold. Parkinson's disease identification utilising voice features has made substantial use of traditional machine learning methods such as Support Vector Machine (SVM), Decision Tree, Naive Bayes, and Logistic Regression.

6.2. PROPOSED TECHNIQUE:

K-Nearest Neighbours is an instance-based, supervised machine learning technique for tasks involving regression and classification. It operates by locating the k nearest data points (neighbours) in the feature space, then using a majority vote among these neighbours to assign the class label. The selection of k, the distance metric, and the quality of the input characteristics all have a significant impact on KNN performance. Because KNN is non-parametric, it can handle complex and non-linear datasets without assuming any previous distribution about the data. Using voice characteristics taken from speech recordings, KNN is used in this study to categorise patients. A feature vector is created from each patient's voice data, and this vector is compared to those in the training dataset in order to conduct classification. To prevent bias in distance computations, feature scaling methods such as Min-Max normalisation are used. The algorithm makes accurate predictions and successfully separates Parkinson's sufferers from healthy people by adjusting hyperparameters like k and distance functions. Because of this, KNN is a straightforward but effective method for clinical decision support in the early identification of Parkinson's disease.



7. DISCUSSION AND CONCLUSION

This study uses speech data to show how machine learning can help detect Parkinson's disease (PD) early. The system offers a straightforward yet effective classification framework by utilising the K-Nearest Neighbours (KNN) method. The method successfully detects early signs of Parkinson's disease (PD), such as jitter, shimmer, and pitch fluctuations. The dataset is made consistent and optimised for trustworthy analysis by normalisation and dimensionality reduction. High accuracy and interpretability are guaranteed via hyperparameter adjustment. The study tackles issues like inadequate transparency in conventional diagnostic techniques and class imbalance. According to their findings, KNN can effectively differentiate between PD patients and healthy people. In addition to traditional

International Journal of Engineering Research and Science & Technology

neurological evaluations, the model provides clinicians with a helpful diagnostic tool. It gives academics a starting point for additional computational healthcare research. The system can handle bigger datasets in the future and is flexible and scalable. Crucially, it fills the knowledge gap between data-driven insights and medical expertise. The technology increases interpretability and dependability by providing probability-based results. The study emphasises speech analysis's potential as a non-invasive diagnostic technique. All things considered, the initiative uses useful machine learning tools to modernise healthcare. It highlights how AI can be used to develop early detection systems that enhance patient outcomes.

8. FUTURE ENHANCEMENT

In the future, this project could be improved by using deep learning models for more intricate speech pattern detection, like CNNs and RNNs. Applications for real-time speech monitoring could be created to help with ongoing patient monitoring. Patients and professionals may find the system easier to use if it is integrated with wearable technology and smartphone apps. Generalisability would be enhanced by broadening the dataset to include more languages and populations. Accuracy could be further increased by feature engineering with sophisticated signal processing methods. Robustness may be improved by hybrid models that combine KNN with additional classifiers. Scalable and remote diagnostic support would be possible with cloud-based deployment. Models could be tailored to each patient's speech baseline to create personalised patient monitoring systems. Early detection may be enhanced by incorporating multimodal data, such as handwriting and gait. Overall, further development will improve the system's comprehensiveness, usability, and therapeutic impact.

9. RESULTS AND DISCUSSION

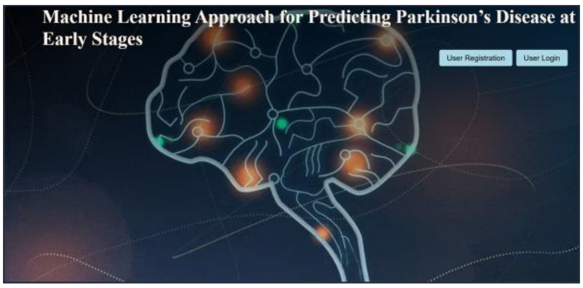


Fig Main Page

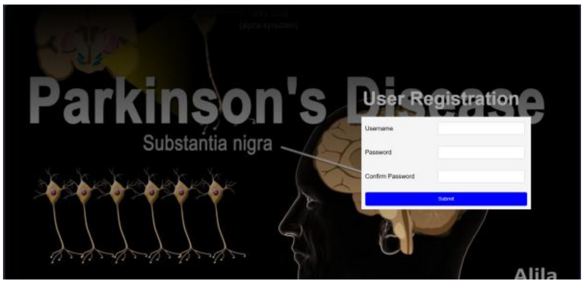


Fig User Registration Page

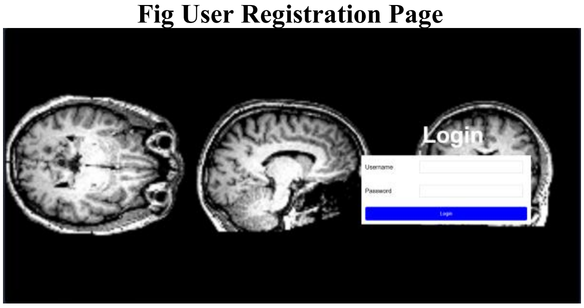


Fig Login Page

Fig. Feature Inputs

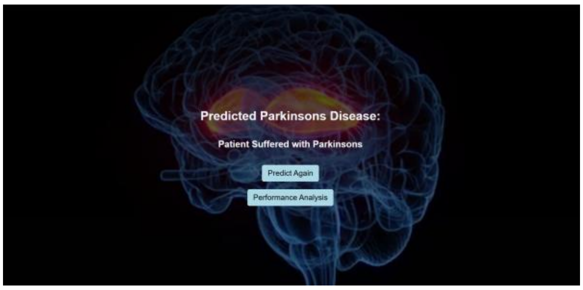
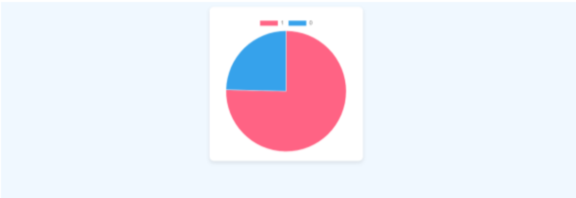


Fig. Results



Performance Analysis

REFERENCES

[l] S. L. Oh, Y. Hagiwara, U. Raghavendra, R. Yuvaraj, N. Arunkumar, M. Murugappan, and U. R. Acharya, "A deep learning approach for Parkinson 's disease diagnosis from EEG signals," Neural Comput. Appl., vol. 32, no. 15, pp. 10927-10933, Aug. 2020.

[2] C. Loconsole, G. D. Cascarano, A. Brunetti, G. F. Trotta, G. Losavio, V. Bevilacqua, and E. Di Sciascio, "A model-free technique based on computer vision and sEMG for classification in Parkinson's disease by using computer-assisted handwriting analysis," Pattern Recognit. Lett., vol. 121, pp. 28-36, Apr. 2019.

[3] b. F. Ertugl-ul, Y. Kaya, R. Tekin, and M. N. Almah, "Detection of Parkinson's disease by shifted one-dimensional local binary patterns from gait," Expert Syst. Appl., vol. 56, pp. 156-163, Sep. 2016.

[4] R. Gupta, M. Khari, D. Gupta, and R. G. Crespo, “Fingerprint image enhancement and reconstruction using the orientation and phase reconstruction," Inf Sci., vol. 530, pp. 201-218, Aug. 2020.

[5] H. M. R. Afzal, S. Luo, M. K. Afzal, G. Chaudhary, M. Khari, and S. A. P. Kumar, "3D face reconstruction from single 2D image using distinctive features," IEEE Access, vol. 8, pp. 180681-180689, 2020.

[6] R. Raj, P. Rajiv, P. Kumar, M. Khari, E. Verdu, R. G. Crespo, and G. Manogaran, "Feature based video stabilization based on boosted Haar cascade and representative point matching algorithm," Image Vis. Comput., vol. 101, Sep. 2020, Art. no. 103957.

[7] R. Gupta, M. Khari, V. Gupta, E. Verdu, X. Wu, E. Herrera- Viedma, and R. G. Crespo, "Fast single image haze removal method for inhomogeneous environment using variable scattering coefficient," Comput. Model. Eng. Sci., vol. 123, no. 3, pp. 1175-1192, 2020.

[8] A. Ma, K. K. Lau, and D. Thyagarajan, "Voice changes in Parkinson's disease: What are they telling us?" J. Clin. Neurosci., vol. 72, pp. 1-7, Feb. 2020.

[9] S. Saravanan, K. Ramkumar, K. Adalarasu, V. Sivanandam, S. R. Kumar, S. Stalin, and R. Amirtharajan, "A systematic review of artificial intelligence (Al) based approaches for the diagnosis of Parkinson's disease," Arch. Comput. Methods Eng., vol. 29, no. 6, pp. 3639-3653, Oct. 2022.

[10] K. A. Shastry, "An ensemble nearest neighbor boosting technique for prediction of Parkinson's disease," H ealthcare Anal., vol. 3, Nov. 2023, Art. no. 100181.

[ll] A. M. Ali, F. Salim, and F. Saeed, "Parkinson's disease detection using filter feature selection and a genetic algorithm with ensemble learning," Diagnostics, vol. 13, no. 17, p. 2816, Aug. 2023.

[12] S. N. H. Bukhari and K. A. Ogudo, "Ensemble machine learning approach for Parkinson's disease detection using speech signals," Mathematics, vol. 12, no. 10, p. 1575, May 2024.

[13] K. Shyamala and T. M. Navamani, "Design of an efficient prediction model for early Parkinson's disease diagnosis," IEEE Access, vol. 12, pp. 137295-137309, 2024.

[14] World Health Org. (2024). Parkinson Disease. Accessed: Mar. 7, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/parkinsondisease

[15] W. Wang, J. Lee, F. Harrou, and Y. Sun, "Early detection of Parkinson's disease using deep learning and machine learning," IEEE Access, vol. 8, pp. 147635-147646, 2020.

[16] M. Hoq, M. N. Uddin, and S.-B. Park, "Vocal feature extraction-based artificial intelligent model for Parkinson's disease detection," Diagnostics, vol. 11, no. 6, p. 1076, Jun. 2021.

[17] Z. Soumaya, B. D. Taoufiq, N. Benayad, K. Yunus, and A. Abdelkrim, "The detection of Parkinson disease using the genetic algorithm and SVM classifier," Appl. Acoust., vol. 171, Jan. 2021, Art. no. 107528.

[18] Y. Liu, Y. Li, X. Tan, P. Wang, and Y. Zhang, "Local discriminant preservation projection embedded ensemble learning based dimensionality reduction of speech data of Parkinson's disease," Biomed. Signal Process. Control, vol. 63, Jan. 2021, Art. no. 102165.

[19] J. Goyal, P. Khandnor, and T. C. Aseri, "A comparative analysis of machine learning classifiers for dysphonia-based classification of Parkinson's disease," Int. J. Data Sci. Analytics, vol. 11, no. l, pp. 69-83, Jan. 2021.

[20] J. Dhar, "An adaptive intelligent diagnostic system to predict early stage of Parkinson's disease using two-stage dimension reduction with genetically optimized LightGBM algorithm," Neural Comput. Appl., vol. 34, no. 6, pp. 4567-4593, Mar. 2022.

[21] Y. Liu, Z. Liu, X. Luo, and H. Zhao, "Diagnosis of Parkinson 's disease based on SHAP value feature selection," Biocybern. Biomed. Eng., vol. 42, no. 3, pp. 856-869, Jul. 2022.

[22] R. Lamba, T. Gulati, and A. Jain, "A hybrid feature selection approach for Parkinson's detection based on mutual information gain and recursive feature elimination," Arabian J. Sci. Eng., vol. 47, no. 8, pp. 10263-10276, Aug. 2022.

[23] S. Hawi, J. Alhozami, R. AlQahtani, D. AlSafran, M. Alqarni, and L. E. Sahmarany, "Automatic Parkinson's disease detection based on the combination of long-term acoustic features and Mel frequency cepstral coefficients (MFCC)," Biomed. Signal Process. Control, vol. 78, Sep. 2022, Art. no. 104013.

[24] K. M. Alalayah, E. M. Senan, H. F. Atlam, I. A. Ahmed, and H. S. A. Shatnawi, "Automatic and early detection of Parkinson's disease by analyzing acoustic signals using classification algorithms based on recursive feature elimination method," Diagnostics, vol. 13, no. 11, p. 1924, May 2023.

[25] A. Iyer, A. Kemp, Y. Rahmatallah, L. Pillai, A. Glover, F. Prior, L. Larson-Prior, and T. Virmani, "A machine learning method to process voice samples for identification of Parkinson's disease," Sci. Reports, vol. 13, no. 1, p. 20615, 2023.

[26] S. Rahman, M. Hasan, A. K. Sarkar, and F. Khan, "Classification of Parkinson's disease using speech signal with machine learning and deep learning approaches," Eur. J. Electr. Eng. Comput. Sci., vol. 7, no. 2, pp. 20--27, Mar. 2023.

[27] R. Alshammri, G. Alharbi, E. Alharbi, and I. Almubark, "Machine learning approaches to identify Parkinson's disease using voice signal features," Frontiers Artif. lntell., vol. 6, Mar. 2023, Art. no. 1084001.

[28] R. Saxena and J. Andrew, "Parkinson's disease identification from speech signals using machine learning models," in Proc. Int. Conj Artif. Intel/. Textile Apparel. Cham, Switzerland: Springer, 2023, pp. 201-213.

[29] P. K. Chawla, M. S. Nair, D. G. Malkhede, H. Y. Patil, S. K. Jindal, A. Chandra, and M. A. Gawas, "Parkinson's disease classification using nature inspired feature selection and recursive feature elimination," Multimedia Tools Appl., vol. 83, no. 12, pp. 35197-35220, Sep. 2023.

[30] N. Singh and P. Tripathi, "An ensemble technique to predict Parkinson's disease using machine learning algorithms," Speech Commun., vol. 159, Apr. 2024, Art. no. 103067.

[31] M. A. Hossain and F. Amenta, "Machine learning-based classification of Parkinson's disease patients using speech biomarkers," J. Parkinson 's Disease, pp. 1-15, 2024.

[32] D. Dua and C. Graff. (2019). UCI Machine Learning Repository: Parkinson 's Disease Classification Data Set. Accessed: Nov. 10, 2024. [Online]. Available: https://archive.ics.uci.edu/dataset/ 470/parkinson+s+disease+classification

[33] A. I. Schein and L. H. Ungar, "Active learning for logistic regression: An evaluation," Mach. Learn., vol. 68, no. 3, pp. 235-265, Aug. 2007.

[34] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, "Hybrid techniques for the diagnosis of acute lymphoblastic leukemia based on fusion of CNN features," Diagnostics, vol. 13, no. 6, p. 1026, Mar. 2023.

[35] E. M. Senan, M. E. Jadhav, and A. Kadam, "Classification of PH2 images for early detection of skin diseases," in Proc. 6th Int. Conj Converg. Technol. (12CT), Apr. 2021, pp. 1-7.

[36] E. Wilson and D. W. Tufts, "Multilayer perceptron design algorithm," in Proc. IEEE Workshop Neural Netw. Signal Process., Nov. 1994, pp. 61-68. [37] A. Kolli. (2021). Understanding Catboost: The Gradient Boosting Algorithm for Categorical Data. [Online]. Available: https://aravindkolli. medium .com/understanding-catboost-the-gradient-boosting-algorithm- for-categorical-data- 73ddb200895d

[38] C. E. Rasmussen, "Gaussian processes in machine learning," in Summer School on Machine Learning. Cham, Switzerland: Springer, 2003, pp. 63-71.

[39] E. M. Senan and M. E. Jadhav, "Diagnosis of dermos copy images for the detection of skin lesions using SVM and KNN," in Proc. 3rd Int. Conj Sustain. Comput., SUSCOM. Cham, Switzerland: Springer, Jan. 2022, pp. 125-134.

[40] M. Vakili, M. Ghamsari, and M. Rezaei, "Performance analysis and comparison of machine and deep learning algorithms for IoT data classification," 2020, arXiv:2001.09636.

[41] M. Shivakoti, S. C. Medaramatla, D. Godavarthi, and N. Shivakoti, "Prog- noza: Parkinson's disease prediction using classification algorithms," EAi Endorsed Trans. Pervasive Health Technol., vol. 9, pp. 1-8, Aug. 2023.