PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING TECHNIQUES

Biradar Vaishnavi[1*], Paramjeet Singh[2], Vaishnavi Biradar[3], Margesh keskar[4]

[1,2,3,4]*Department of CSE (Data Science), Guru Nanak Dev Engineering Collage Bidar, Karnataka , India*

*biradarvaishnavi75@gmail.com

Abstract

Parkinson's disease is a long-term neurological disorder that gradually affects both movement and non-movement functions in humans. Detecting this disease at an early stage is difficult because symptoms appear slowly over time. Recently, machine learning has become an effective approach for analyzing complex medical data. In this work, a machine learning-based system is developed to identify Parkinson's disease using voice-related biomedical data. The dataset includes important features such as jitter, shimmer, and frequency parameters. Various preprocessing steps like cleaning, handling missing values, scaling, and feature selection are applied to improve the dataset quality. Two supervised learning algorithms, Random Forest and XGBoost, are used for classification. The results show that the model achieves high accuracy in distinguishing between healthy individuals and patients. This system can support doctors in early diagnosis and better decision-making.

Keywords: Parkinson's Disease, Machine Learning, Random Forest, XGBoost, Voice Signal Analysis, Predictive Modeling

1. Introduction

Parkinson's disease is one of the most common neurological disorders that mainly impacts the motor system. It leads to symptoms such as tremors, stiffness, slow movement, and difficulty in maintaining balance. Apart from these, patients may also experience non-motor issues like depression, memory problems, and speech difficulties.

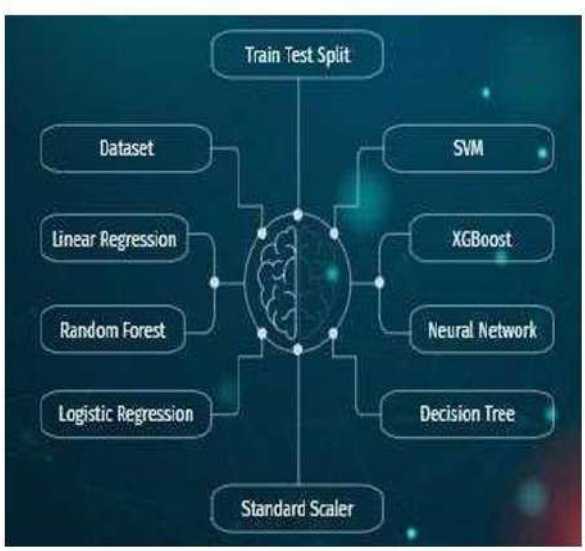
Traditional diagnosis methods rely heavily on clinical observations, which may delay detection because symptoms develop gradually. Therefore, there is a need for automated systems that can identify the disease at an early stage.

Machine learning provides powerful tools to analyze medical data and identify patterns. This study focuses on using voice-based features to predict Parkinson's disease since speech changes are one of the early indicators. The goal is to build a model that can accurately classify whether a person is healthy or affected.

2. Related Work

Many researchers have worked on detecting Parkinson's disease using machine learning techniques. Different types of data such as gait patterns, speech signals, and handwriting samples have been used to study the disease.

Earlier studies using gait analysis achieved around 85–87% accuracy. Later, machine learning algorithms like KNN, Decision Trees, and Random Forest improved performance up to around 90% or more.



Voice-based detection methods are gaining popularity because they are simple and non-invasive. Some studies have reported accuracy above 94% using speech features. Deep learning models such as CNNs have further increased accuracy levels up to 96–98%.

However, challenges like data imbalance, irrelevant features, high computation cost, and lack of interpretability still exist. This work attempts to overcome these issues using proper preprocessing and efficient algorithms.

3. Materials and Methods

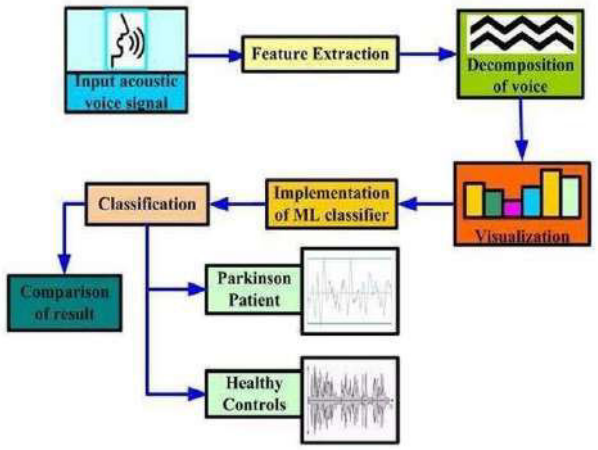
3.1 Dataset Description

The dataset used in this study is taken from the UCI Machine Learning Repository. It contains 195 records with 23 attributes related to voice measurements.

Each record is labeled as:

- **0 → Healthy person**
- **1 → Parkinson's patient**

The dataset includes important voice features such as jitter, shimmer, and frequency values.

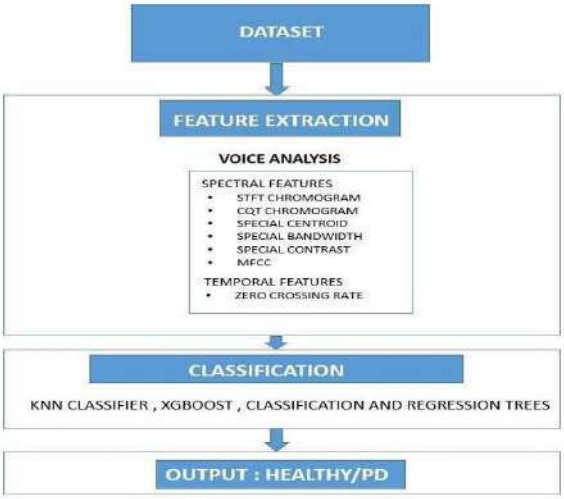


3.2 System Architecture

The system follows a step-by-step process:

1. Input voice data
2. Data preprocessing
3. Feature selection
4. Model training
5. Prediction output

This structured workflow ensures better performance and accuracy.



3.3 Data Preprocessing

Data preprocessing is an important step to improve model quality. The following steps are performed:

- **Data Cleaning:** Removing noise and duplicate values
- **Handling Missing Values:** Filling missing data using suitable methods
- **Feature Engineering:** Improving feature usefulness
- **Feature Scaling:** Normalizing values to avoid bias

These steps help in improving model efficiency and accuracy.

3.4 Dataset Splitting

The dataset is divided into:

- **Training Data →** Used to train the model
- **Testing Data →** Used to evaluate performance

This helps in checking how well the model works on new data.

3.5 Feature Selection

Only important features are selected to improve performance.

Key features include:

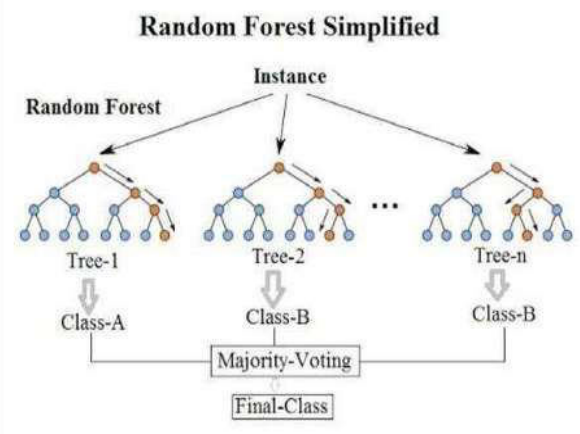
- Jitter
- Shimmer
- Frequency parameters

These features are strongly related to voice abnormalities in Parkinson's patients.

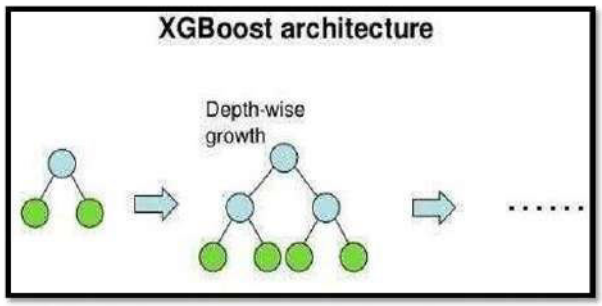
3.6 Machine Learning Algorithms

Two algorithms are used:

- **Random Forest:** Combines multiple decision trees to improve accuracy and reduce overfitting.



XGBoost:
A powerful boosting algorithm that gives high performance and faster computation.



These algorithms are trained using the processed training dataset to learn patterns associated with Parkinson's disease.

3.7 Model Training and Evaluation

The models are trained using training data and evaluated using testing data.

Performance is measured using:

- Accuracy
- Precision
- Recall
- F1-score

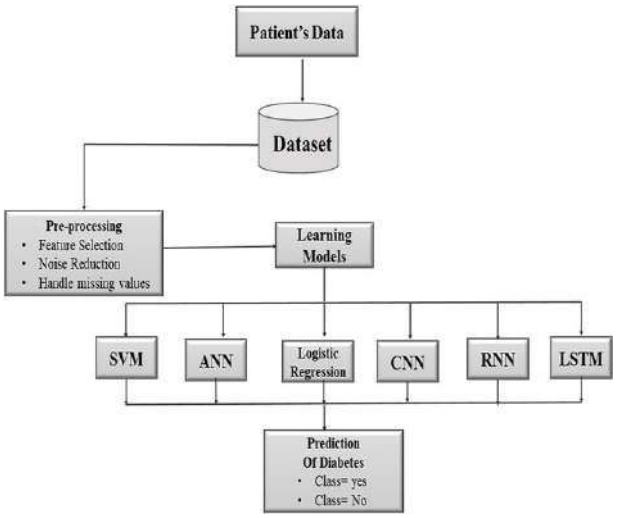
These metrics help in understanding how well the model is performing.

3.8 Prediction Output

The model gives output as:

- **0 → No disease**
- **1 → Parkinson's disease present**

This helps in early detection and medical support.

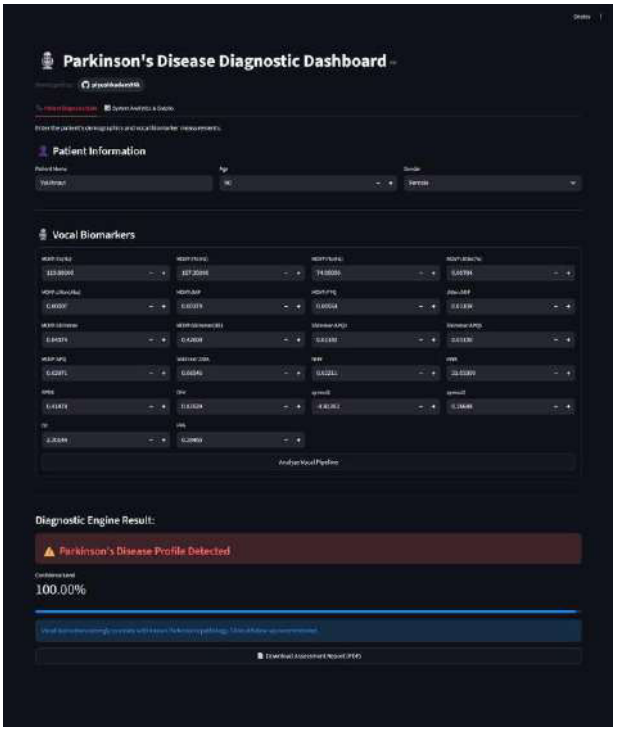


4. Results and Discussion

The models are evaluated using standard metrics and confusion matrix analysis.

- Random Forest provides stable and reliable results due to its ensemble nature.
- XGBoost gives higher accuracy because of its advanced boosting technique.

Voice features like jitter and shimmer are found to be strong indicators of the disease.



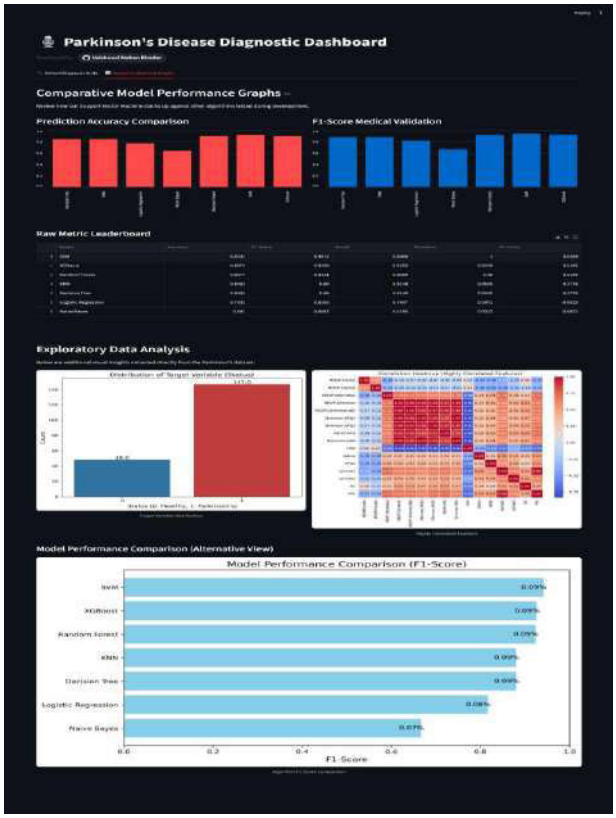
Advantages

- Early detection of disease
- Non-invasive method
- Cost-effective
- Helps doctors in decision-making

Limitations

- Depends on dataset quality
- May not generalize well to all populations

Needs real-world testing



5. Conclusion

This study presents a machine learning-based system for detecting Parkinson's disease using voice data. The use of Random Forest and XGBoost improves prediction accuracy significantly.

The system shows strong potential in assisting doctors for early diagnosis. Machine learning can play a major role in improving healthcare systems and patient outcomes.

Future Scope (Rewritten)

Future improvements can include:

- Using deep learning models
- Combining multiple data types (voice + gait + handwriting)
- Developing mobile/web applications for real-time detection

6.References

[1] M. A. Nair and V. K. Menon, "Parkinson's Disease Diagnosis using Machine Learning: A Review," 2021.

[2] J. R. Silva et al., "Machine Learning Approaches for Parkinson's Disease Diagnosis using Gait Analysis," 2020.

[3] B. Singh and R. Singh, "Comparative Study of Machine Learning Algorithms for Parkinson's Disease Diagnosis," 2019.

[4] N. F. Reis et al., "Parkinson's Disease Detection using Voice Analysis and Machine Learning," 2020.

[5] A. F. Ahmed et al., "Deep Learning-Based Parkinson's Disease Detection," 2020.

[6] Z. A. H. Shah et al., "Handwritten Spiral Analysis for Parkinson's Disease Diagnosis," 2020.

[7] Meireles, J. and Massano, J., 2012. Cognitive impairment and dementia in Parkinson's disease: clinical features, diagnosis, and management. Frontiers in neurology, 3, p.88.

[8]. Samii A., J.G. Nutt, B.R. Ransom. Parkinson's disease. Lancet, 363 (2004), pp. 1783- 1793.

[9]. Khan Academy. (2015). Movement signs and symptoms of Parkinson's disease | NCLEX- RN | Khan Academy.

[10]. Litvan, I., Goldman, J.G., Tröster, A.I., Schmand, B.A., Weintraub, D., Petersen, R.C. Mollenhauer, B., Adler, C.H., Marder, K., Williams-Gray, C.H. and Aarsland, Movement Disorder Society Task Force guidelines. Movement disorders, 27(3), pp.349356.

[11]. Berg, D., Lang, A.E., Postuma, R.B., Maetzler, W., Deuschl, G., Gasser, T., Siderowf, A., Schapira, A.H., Oertel, W., Obeso, J.A. and Olanow, C.W., 2013. Changing the research criteria for the diagnosis of Parkinson's disease: obstacles and opportunities. The Lancet Neurology, 12(5), pp.514-524.

[12]. R. Lafuente, J. M. Belda, J. Sanchez-Lacuesta, C. Soler, and J. Prat, "Design and test of neural networks and statistical classifiers in computer-aided movement analysis: a case study on gait analysis," 1997.

[13]. J. G. Barton and A. Lees, "An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams," 1997.

[14]. R. Begg and J. Kamruzzaman, "Neural networks for detection and classification of walking pattern changes due to ageing," 2006.

[15]. I. Rustempasic and M. Can, "Diagnosis of Parkinson's Disease using Fuzzy C Means Clustering and Pattern Recognition," Southeast Europe Journal of Soft Computing, vol. 2, no. 1, Mar. 2013, doi: 10.21533/scjournal. v2i1.44.

[16]. R. Geetha, R. Professor, & Head, and G. Sivagami, "Parkinson Disease Classification using Data Mining Algorithms," 2011.