



DETECTION OF PARKINSON'S DISEASE USING MACHINE LEARNING

Dr.R.Hemalatha, R.Blessy

Head & Associate Professor, Research Scholar
PG & Research Department of Computer Science
Tiruppur Kumaran College for Women, Tirupur, India

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative disorder with motor and non-motor symptoms. It is a degenerative movement illness, its symptoms begin mildly and progressively worsen. Parkinson's disease usually starts with mild hand or finger tremors before developing into other symptoms over time. Parkinson's disease is a neurological condition that impairs balance, muscle function, and movement. Although the illness can strike anyone at any age, it most commonly strikes those between the ages of 55 and 75. Accurate early detection is crucial for patient care and clinical trials. Machine learning (ML) methods allow non-invasive analysis of digital biomarkers. Early detection is essential for managing sickness effectively. The K-Nearest Neighbor (KNN) and Feed-forward Neural Network (FNN) models are two Machine Learning (ML) and Deep Learning (DL) techniques that we utilize in this study to differentiate between individuals with Parkinson's disease (PD) and healthy individuals based on voice signal characteristics.

Keywords :

Parkinson's Disease, Machine Learning, K-Nearest Neighbor (KNN), Feed-forward Neural Network (FNN), Deep Learning (DL)

I. INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized primarily by motor symptoms (tremor, bradykinesia, rigidity, postural instability) and a range of non-motor signs. The most prevalent sign of Parkinson's disease is tremor. Patients usually experience a sporadic rhythmic tremor that begins in one finger and gradually spreads to the entire arm. When an arm or leg is at rest or in an unsupported position, some individuals experience tremors. One or both sides of the body may experience these tremors. Tremors can impact the head, lips, tongue, and feet if treatment is not received. Clinical diagnosis typically occurs after motor symptoms appear, by which time substantial neuronal loss has occurred. There is growing interest in ML methods that exploit digital biomarkers from voice, smartphone sensors, wearable devices, handwriting, and medical imaging to enable earlier, objective detection and monitoring. Recent reviews show

rapid growth in ML/DL applications for PD detection and progression monitoring, and large mobile datasets (e.g., mPower) are widely used for method development. As symptoms worsen over time, Parkinson's disease advances.

The progression can be divided into five stages:

Stage 1: Only minor symptoms, such as tremor, are felt on one side of the body. Changes in posture, facial expressions, and walking may start to occur. Symptoms often do not interfere with daily activities.

Stage 2: Movement issues, tremors, and stiffness affect both sides of the body. Walking and postural issues become more apparent. Living alone is completely possible, although it gets increasingly difficult to perform daily duties.

Stage 3: Balance issues and an increase in falls occur. The tempo decreases. Independent living is still achievable even though eating, dressing, and taking a shower may require assistance.

Stage 4: Parkinson's disease symptoms have gotten much worse and are seriously interfering with daily activities. You need assistance walking, such as a walker. Living on your own is impossible.

Stage 5: The most severe symptoms can make walking impossible. At the most incapacitating time, some people end up bedridden. 24/7 nursing care is required. Mental disability could manifest as anything from delusions to hallucinations.

II. PROBLEM IDENTIFICATION :

- Parkinson Disease is a progressive neurological condition that is not as easy to detect at the early-stage because the symptoms are subtle and similar to other neurological diseases.
- Dysphonia is noted in most patients and early indicator, but the speech-related features are underutilized in accurate prediction.
- Current machine learning systems can fail to make consistent predictions with image data due to the ambiguity of PD symptoms with regard to structural images.
- The majority of the available Parkinson algorithms easily are prone to class imbalance, which contributes to a biased forecast and low model robustness.
- The high-dimensionality of speech data creates superfluous features and effective feature selection is required to make models more accurate and at the same time efficient.
- Most current works are limited to either or disease outcome and neglect disease-staging and subtype-specific considerations that are important in monitoring progression and clinical management.

III. PROPOSED METHODOLOGY

The main goal is to discriminate between individuals with Parkinson's disease and healthy individuals using distinctive traits extracted from speech signals. The potential outcome is to increase diagnostic precision. We use a number of techniques to maximize model performance, such as the Synthetic Minority Over-sampling Technique (SMOTE) to solve class imbalance, Feature Selection to find the most pertinent features, and RandomizedSearch for hyperparameter tuning, with the help of sophisticated methods like Recursive Feature

Elimination and Feature Importance in the critical feature selection assignment, the selected diagnostic methodology effortlessly blends classification and selection procedures.

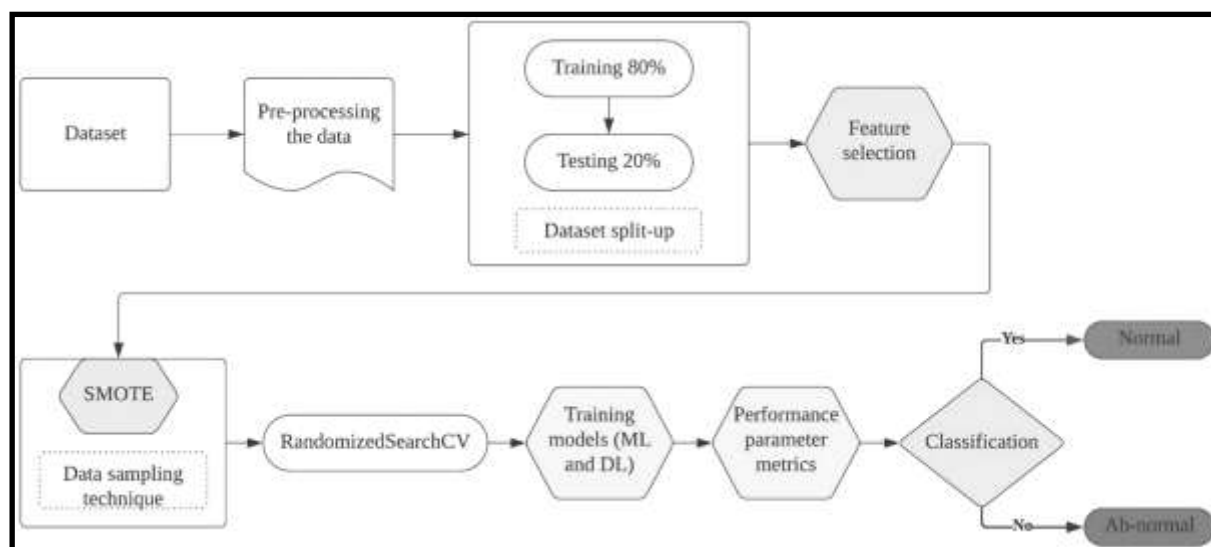


Fig : Proposed model architecture.

(i) Data Collection: The dataset used in this study consists of voice recordings from individuals with and without Parkinson's disease. By transforming these recordings into structured CSV format, a range of vocal attributes, such as pitch, jitter, shimmer, and harmonic-to-noise ratio, have been recorded.

(ii) Preprocessing: To address the class imbalance in the dataset, we employed the Synthetic Minority Over-sampling Technique (SMOTE). This approach ensures a balanced mix of Parkinson's and healthy patients in the training set by generating false samples for the minority class.

(iii) Feature Selection: We employed Recursive Feature Elimination (RFE) to select the most relevant speech traits for Parkinson's disease classification. RFE iteratively removes the least significant features based on the performance of a Support Vector Machine (SVM) model until a subset of traits most pertinent to the classification objective is selected.

(iv) KNN: This model classifies people based on the majority class of their nearest neighbors in the feature space.

(v) FNN: This neural network consists of an input layer, one or more hidden layers, and an output layer. We optimized the network configuration and architecture using Randomized Search

(vi) Hyperparameter Tuning: We performed hyperparameter tuning for both models using RandomizedSearchCV, which randomly tests a large number of parameter combinations to determine the optimal settings that maximize model performance.

(vii) Evaluation: The models' performance was evaluated using accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of the models' ability to reliably differentiate between individuals with and without Parkinson's disease.

IV. PREPROCESSING

- **Voice:** resample to 16 kHz, apply pre-emphasis, voice activity detection, remove low-quality clips. Extract MFCCs, formant frequencies, jitter, shimmer, Harmonics-to-Noise Ratio (HNR), and compute short-time Fourier transform (STFT) for spectrograms. Prior studies use these features successfully for PD classification.

- **Tapping/Gait:** low-pass filter, sensor alignment, normalization by participant baseline, compute gait parameters (step time variance, stride length proxies) and tapping inter-tap intervals and statistics.
- **Data balancing:** account for class imbalance via stratified sampling, SMOTE variants only on training folds (avoid leakage).
- **Dataset biases and label noise:** self-reported diagnoses and convenience sampling (e.g., mPower) can introduce noise and demographic bias.
- **Overfitting and validation leakage:** many published works use sample-level CV rather than participant-level splits, inflating reported accuracy. Rigorous participant-level validation and external holdouts are required.
- **Clinical integration:** a positive screening test must be clinically actionable; early detection without available disease-modifying therapies raises ethical issues. Recent work on blood-based biomarkers and wearables suggests complementary clinical pathways but highlights need for validation.

V. CONCLUSION

Machine learning offers promising tools for PD detection using non-invasive digital biomarkers. Diagnosing Parkinson's disease (PD) uses cutting-edge deep learning and machine learning algorithms to outperform existing approaches in terms of accuracy. Each model improves our knowledge of how to diagnose Parkinson's disease, highlighting the versatility and potency of several machine learning approaches in addressing this significant medical problem. Feed-forward neural networks (FNN) and K-Nearest Neighbors (KNN) are used to thoroughly evaluate their effects on model performance. The dataset is carefully divided into two subsets: 80% for training and 20% for testing and validation in order to give reliable evaluation. The primary advantage of our approach is its ability to enable prompt and precise PD detection, which is crucial for improving patient outcomes.

V. REFERENCES

1. Alshammri, R., Alharbi, G., Alharbi, E. & Almubark, I. Machine learning approaches to identify Parkinson's disease using voice signal features. *Front. Artif. Intell.* 6(1), 1–8 (2023).
2. Cherukuvada, S. & Kayalvizhi, R. Feature selection with deep belief network for epileptic seizure detection on EEG signals. *Comput. Mater. Contin.* 75(2), 4101–4118 (2024).
3. Martinez-Eguiluz, M. *et al.* Diagnostic classification of Parkinson's disease based on non-motor manifestations and machine learning strategies. *Neural Comput. Appl.* 35(1), 5603–5617 (2023).
4. Tabashum T. *et al.*, *Machine Learning Models for Parkinson Disease: Systematic Review*, JMIR Med Inform, 2024.
5. Srinivasan, S. *et al.* Detection and classification of adult epilepsy using hybrid deep learning approach. *Sci. Rep.* 13(1), 1–17 (2023).