Life Science Informatics Publications
Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences
Journal Home page http://www.rjlbpcs.com/
RJLBPCS
ISSN 2454-6348

Original Research Article **DOI: 10.26479/2025.1202.01**

MACHINE LEARNING APPROACHES FOR THE EARLY DIAGNOSIS OF PARKINSON'S DISEASE

A. Kirubha, D. Umanandhini

Department of Computer Science, Kovai Kalimagal College of Arts and Science, Coimbatore – 641109.

ABSTRACT: Parkinson's Disease (PD) is a progressive neurological disorder that affects body movement, speech, and coordination. Early detection of the disease is important, as it helps doctors provide timely treatment and manage symptoms more effectively. In recent years, machine learning techniques have been widely used to assist in the early detection of Parkinson's disease using different types of medical data, such as voice recordings, spiral drawing tests, and clinical measurements. The machine learning algorithms, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), Neural Networks, and Convolutional Neural Networks (CNN) have been used to predict Parkinson's disease. This paper examines and compares several research works that use machine learning methods for Parkinson's disease detection. The analysis focuses on commonly used datasets, preprocessing techniques, and classification algorithms used for Parkinson's disease prediction. The results show that machine learning models can achieve high prediction accuracy and can support healthcare professionals in the early diagnosis of Parkinson's disease. However, challenges such as small datasets, data imbalance, and limited real-world clinical implementation exist in this disease prediction. Therefore, further research is needed to develop more reliable and practical models for Parkinson's disease detection.

Keywords: CNN, Early Diagnosis, KNN, Parkinson's Disease.

Article History: Received: March 20, 2026; Revised: March 28, 2026; Accepted: April 10, 2026.

Corresponding Author: A. Kirubha*

Department of Computer Science, Kovai Kalimagal College of Arts and Science, Coimbatore – 641109, Tamil Nadu, India. Email Address: kirubha@kkcas.edu.in

1. INTRODUCTION

Parkinson's disease (PD) is one of the most common neurodegenerative disorders that mainly affects

the elderly population. It occurs due to the gradual loss of dopamine-producing neurons in a region of the brain called the substantia nigra. Dopamine is a chemical that plays an important role in controlling body movements[1]. When dopamine levels decrease, communication between nerve cells becomes weaker, which leads to problems in movement control. Patients with Parkinson's disease commonly experience symptoms such as tremors (shaking of hands or limbs), muscle stiffness, slow movement (bradykinesia), balance difficulties, and speech problems. These symptoms start slowly and worsen over time as the disease progresses[2]. According to the World Health Organization (WHO), more than 8.5 million people worldwide were living with Parkinson's disease in 2019. The prevalence of Parkinson's disease increases with age, and studies indicate that about 1–2% of people above the age of 60 are affected by PD. Recent global health studies propose that the number of Parkinson's disease cases has more than doubled over the past 25 years, mainly due to population aging and increased life expectancy. Traditional diagnosis of Parkinson's disease mainly relies on clinical observation and neurological examination. Doctors evaluate physical symptoms, patient history, and motor function tests to confirm the disease. The early symptoms of Parkinson's disease are minor and resemble other neurological conditions. Therefore, the early diagnosis becomes difficult, and the disease is identified only after significant progression [3]. In recent years, advances in Artificial intelligence (AI) and machine learning (ML**)** have created new opportunities for improving medical diagnosis. Machine learning models can analyze large volumes of medical data and identify hidden patterns that cannot be recognized by human experts. For Parkinson's disease detection, researchers have used different types of data, such as voice recordings, spiral drawing tests, handwriting samples, and clinical measurements, to identify early signs of the disease. The researchers applied machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Deep learning models to classify individuals as either healthy or affected by Parkinson's disease[4]. These approaches help in developing automated and non-invasive diagnostic systems that support healthcare professionals in early detection. This paper analyzes several experimental studies that use machine learning techniques for Parkinson's disease detection. The paper compares the datasets, preprocessing methods, and machine learning algorithms used in these studies and highlights the strengths and limitations of the existing approaches.

2. MATERIALS AND METHODS

2.1 Datasets

The research works that are discussed in this review use publicly available datasets for Parkinson's disease detection. These datasets mainly contain biomedical voice measurements, handwriting patterns, or clinical features collected from both Parkinson's patients and healthy individuals. Using standardized datasets helps researchers evaluate and compare the performance of different machine learning algorithms. One of the most widely used datasets is the UCI Parkinson's dataset, which is

available in the UCI Machine Learning Repository. This dataset contains voice recordings collected from individuals with Parkinson's disease and from healthy persons. The dataset includes several acoustic features such as jitter, shimmer, and fundamental frequency measures that represent variations in voice signals [5]. These voice characteristics are important because Parkinson's disease often affects vocal cord control, leading to noticeable changes in speech patterns. Another commonly used dataset is the MDVP (Multi-Dimensional Voice Program) dataset, which provides detailed voice measurements obtained using specialized medical voice analysis software. This dataset includes multiple parameters related to voice stability and frequency variations, which are useful indicators for identifying Parkinson's disease [6]. Some researchers used spiral drawing datasets to analyze handwriting patterns. In these datasets, individuals are asked to draw spiral shapes using a pen or stylus. Patients with Parkinson's disease often show irregular and distorted spiral patterns due to reduced motor control and tremors. These visual patterns can be analyzed using image processing and machine learning techniques to detect Parkinson's disease.

Table 1: Common Datasets Used for Parkinson's Disease Detection

Dataset Name	Data Type	Number of Instances	Number of Features	Description
UCI Parkinson's Dataset	Voice recordings	195	22	Contains biomedical voice measurements from Parkinson's patients and healthy individuals
MDVP Voice Dataset	Voice signals	~200 samples	Multiple acoustic features	Includes detailed voice measurements using the Multi-Dimensional Voice Program
Spiral Drawing Dataset	Handwriting / Image	Varies across studies	Image-based features	Contains spiral drawings used to analyze motor control abnormalities

These datasets provide valuable information that allows researchers to develop and test machine learning models for the early detection of Parkinson's disease.

2.2 Data Preprocessing

Numerous preprocessing techniques are applied to improve the quality and reliability of the dataset. Preprocessing helps remove noise, handle missing values, balance class distribution, and select relevant features. One important preprocessing technique used is SMOTE (Synthetic Minority Oversampling Technique). Medical datasets suffer from class imbalance, where the number of samples in one class (for example, healthy individuals) is much larger than the other class (Parkinson's patients). SMOTE addresses this problem by generating synthetic samples for the minority class, which helps create a balanced dataset and improves the performance of machine

learning models. Another important preprocessing step is feature selection. In many biomedical datasets, a large number of features are available, but not all of them contribute equally to the classification process. Feature selection techniques help identify the most relevant attributes while removing redundant or irrelevant features [7]. This reduces computational complexity and improves classification accuracy. Some studies also apply dimensionality reduction techniques, such as Principal Component Analysis (PCA). PCA transforms the original high-dimensional dataset into a smaller set of new variables called principal components while preserving most of the important information. This helps simplify the dataset and reduce noise, making the machine learning model more efficient.

Table 2: Preprocessing Techniques Used in PD Detection Studies

Preprocessing Technique	**Purpose**	**Benefit**
Data Cleaning	Remove noise and missing values	Improves data quality
SMOTE	Balance class distribution	Prevents model bias
Feature Selection	Identify relevant features	Improves accuracy and reduces complexity
PCA	Reduce dimensionality	Simplifies the dataset while preserving important information
Normalization	Scale feature values	Ensures fair comparison between features

These preprocessing steps play an important role in preparing the dataset before applying machine learning algorithms. Proper preprocessing improves the reliability and accuracy of Parkinson's disease detection models.

2.3 Machine Learning Algorithms Used for Parkinson's Disease Detection

Machine learning algorithms play an important role in the early detection of Parkinson's disease by analyzing medical datasets and identifying patterns associated with the disease [8]. Many researchers have applied different machine learning techniques to classify individuals as either healthy or affected by Parkinson's disease. The performance of these algorithms depends on the dataset, preprocessing techniques, and feature selection methods used in the study.One commonly used algorithm is the Support Vector Machine (SVM). SVM is a supervised learning algorithm that works by finding the optimal boundary that separates two classes of data. It is widely used in Parkinson's disease detection because it performs well with high-dimensional datasets and provides good classification accuracy. Another popular algorithm is Random Forest (RF). Random Forest is an ensemble learning method that constructs multiple decision trees and combines their outputs to make a final prediction. This approach improves classification performance and reduces the risk of overfitting. K-Nearest Neighbors (KNN) is also used in several studies. KNN is a simple

classification algorithm that assigns a class label based on the majority class of its nearest neighbors in the dataset[9]. Although it is easy to implement, its performance may decrease when the dataset becomes very large. Artificial Neural Networks (ANN) have also been applied for Parkinson's disease detection. Neural networks simulate the structure of the human brain and are capable of learning complex relationships within data. They are especially useful when dealing with non-linear patterns in biomedical datasets. In recent years, Convolutional Neural Networks (CNN) have gained attention in Parkinson's disease research [10]. CNN models are mainly used for analyzing image-based datasets such as spiral drawings and handwriting samples. These models can automatically extract important features from images and improve detection accuracy.

Table 3: Machine Learning algorithms used in PD detection.

Algorithm	**Description**	**Advantages**	**Limitations**	**Reported Accuracy (Approx.)**
Support Vector Machine (SVM)	Separates data using an optimal hyperplane	Works well with high-dimensional data, high accuracy	Sensitive to parameter selection	85–95%
Random Forest (RF)	Ensemble method using multiple decision trees	Reduces overfitting, good performance	Requires more computational resources	88–96%
K-Nearest Neighbors (KNN)	Classifies based on nearest neighbors	Simple and easy to implement	Slow with large datasets	80–92%
Artificial Neural Network (ANN)	Brain-inspired model with interconnected neurons	Handles complex patterns	Requires more training data	85–94%
Convolutional Neural Network (CNN)	Deep learning model for image analysis	Automatically extracts features from images	Computationally expensive	90–97%

2.4 Comparison of Existing Studies

Parkinson's Disease (PD) is a progressive neurological disorder that affects motor and speech functions, making early diagnosis essential for effective treatment and management. In recent years, machine learning (ML) and deep learning (DL) techniques have been widely applied for the early detection of PD using various types of data such as voice signals, speech recordings, and medical images[11]. Different studies have explored diverse preprocessing methods, feature selection techniques, and classification algorithms to improve detection accuracy and reliability. The

performance of PD detection models varies depending on the dataset used, the nature of the data, and the computational techniques applied. Traditional machine learning models such as Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN), as well as advanced approaches including deep learning and hybrid models, have demonstrated significant potential in identifying PD at early stages[12]. The following table presents a comparative analysis of various existing approaches for Parkinson's Disease detection, highlighting the datasets used, data types, preprocessing techniques, algorithms applied, best-performing models, and their respective performance metrics.

Table 4: Comparison of Existing Methods

S.No	Study / Approach	Dataset Used	Data Type	Preprocessing / Feature Selection	Algorithms Used	Best Model	Performance
1	Early Detection of PD using ML in Telemedicine	MDVP Voice Dataset (30 subjects)	Voice Signals	Feature extraction	SVM, RF, KNN, Logistic Regression	Random Forest	91.83 % accuracy
1	PD Detection using Voice Signals	UCI Parkinson's Dataset (195 recordings)	Voiceignals	SMOTE, Feature Selection	KNN, FNN, KSVM	FNN	99.11 % accuracy
2	PD Detection using Spiral Images	Spiral Drawing Dataset	Image Data	Transfer Learning	DenseNet121, VGG16, InceptionV3, LeNet	DenseNet121	98.44% accuracy
4	Privacy-Preserving PD Detection	PD Speech Datasets	Speech Signals	Multi-Cluster Feature Selection	HKNN, LCCB, SVM, RF	HKNN	95.2% accuracy

S.No	Study / Approach	Dataset Used	Data Type	Preprocessing / Feature Selection	Algorithms Used	Best Model	Performance
5	Speech Signal Based PD Prediction	UCI Dataset (188 subjects)	Voice Signals	SMOTE, PCA, ANOVA	SVM, XGBoost, DNN	DNN	High accuracy
6	Ensemble Learning for PD Detection	Voice Dataset	Voice Signals	Filter Feature Selection, Genetic Selection	Decision Tree, RF, XGBoost	Random Forest	100% accuracy
7	Feature Selection Based PD Detection	Voice Dataset (240 records)	Voice Signals	Filter and Wrapper Methods	KNN	KNN	88.33% accuracy
8	Hybrid Deep Learning Approach	UCI Speech Dataset	Speech Signals	Feature extraction	CNN, CLSTM, SVM	CNN-CLSTM	Improved performance
9	ML-based PD Detection using Biomedical Voice Data	Voice Dataset	Voice Signals	Data preprocessing	SVM, RF, KNN	Random Forest	High accuracy

3. RESULTS AND DISCUSSION

The studies reviewed in this paper demonstrate that machine learning techniques can effectively detect Parkinson's disease using different types of medical data. Voice signal analysis is the most commonly used approach because speech changes often appear in the early stages of the disease. Among the traditional machine learning algorithms, Random Forest, Support Vector Machine, and K-Nearest Neighbors have shown strong performance in classification tasks. Random Forest is

frequently reported as one of the best-performing models due to its ability to combine multiple decision trees and reduce overfitting. Deep learning approaches such as Neural Networks and Convolutional Neural Networks have also achieved high accuracy in recent studies. For example, the Feed Forward Neural Network model achieved an accuracy of more than 99%, while CNN models successfully classified spiral drawings with an accuracy above 98%. Feature selection and dimensionality reduction methods also play an important role in improving prediction accuracy. Techniques such as PCA, ANOVA, and genetic feature selection help remove irrelevant features and improve the efficiency of machine learning models. Although the results are promising, some limitations remain. Many studies use relatively small datasets, which may affect the generalization ability of the models. In addition, several models have not yet been tested in real clinical environments. Future research should focus on using larger datasets, combining multiple types of medical data, and developing models that can be applied in practical healthcare systems.

4. CONCLUSION

Parkinson's disease is a serious neurological disorder that affects millions of people worldwide. Early detection is essential for providing timely treatment and improving the quality of life for patients. Machine learning techniques have shown strong potential in detecting Parkinson's disease using voice recordings, spiral images, and biomedical measurements. Several algorithms, such as Random Forest, Support Vector Machine, K-Nearest Neighbors, and deep learning models, have achieved high classification accuracy in various studies. Despite these encouraging results, further research is required to improve the reliability and practical applicability of machine learning models. Future studies should focus on larger datasets, better validation techniques, and integration of multiple data sources. Machine learning-based diagnostic systems have the potential to become valuable tools in healthcare, supporting doctors in the early detection and monitoring of Parkinson's disease.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals or humans were used for the studies that are based on this research.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

No

REFERENCES

1. Srinivasan S, Parthasarathy P. Detection of Parkinson disease using multiclass machine learning approach. Sci Rep. 2024;14:13813.
2. Govindu A, Palwe S. Early detection of Parkinson's disease using machine learning. Procedia Comput Sci. 2023;218:249–261.
3. Shanthappa PM, Bayari M, Abhilash GB, Gokul KV, Ashish PJ. Parkinson's disease detection using InceptionV3: A deep learning approach. MethodsX
4. Du Q, Shen J, Wen P, Chen X. Parkinson's disease detection by using machine learning method based on local classification on class boundary. Discov Appl Sci. 2024;4.
5. Kannan HK, Pooja MR, Ravi V. Predictive analysis of Parkinson's disease using machine learning. Open Bioinform J. 2025;18.
6. Ali AM, Salim F, Saeed F. Parkinson's disease detection using filter feature selection and a genetic algorithm with ensemble learning. Diagnostics. 2023.
7. Saeed F, Al-Sarem M, Al-Mohaimeed M, Emara A, Boulila W, Alasli M, et al. Enhancing Parkinson's disease prediction using machine learning and feature selection methods. Comput Mater Continua. 2022;73(1):123–138.
8. Vishala G, Krishnan R. A survey on Parkinson's disease detection using hybrid deep learning. Int J Eng Res Technol. 2024;13(5).
9. Tabashum T, Snyder RC, O'Brien MK, Albert MV. Machine learning models for Parkinson disease: Systematic review. JMIR Med Inform. 2022;10(8).
10. Shokrpour S, MoghadamFarid A, Bazzaz Abkenar S, Haghi Kashani M, Akbari M, Sarvizadeh M. Machine learning for Parkinson's disease: A comprehensive review of datasets, algorithms, and challenges. npj Parkinsons Dis. 2025.
11. Mei C, Desrosiers C, Frasnelli J. Machine learning for the diagnosis of Parkinson's disease: A review of literature. Front Aging Neurosci. 2021;13:633752.
12. Laxmikantha LK, PS A, PK L, GP P, GSM S. A survey paper on Parkinson's disease detection using machine learning. Int J Adv Res Comput Commun Eng. 2025;14(6).