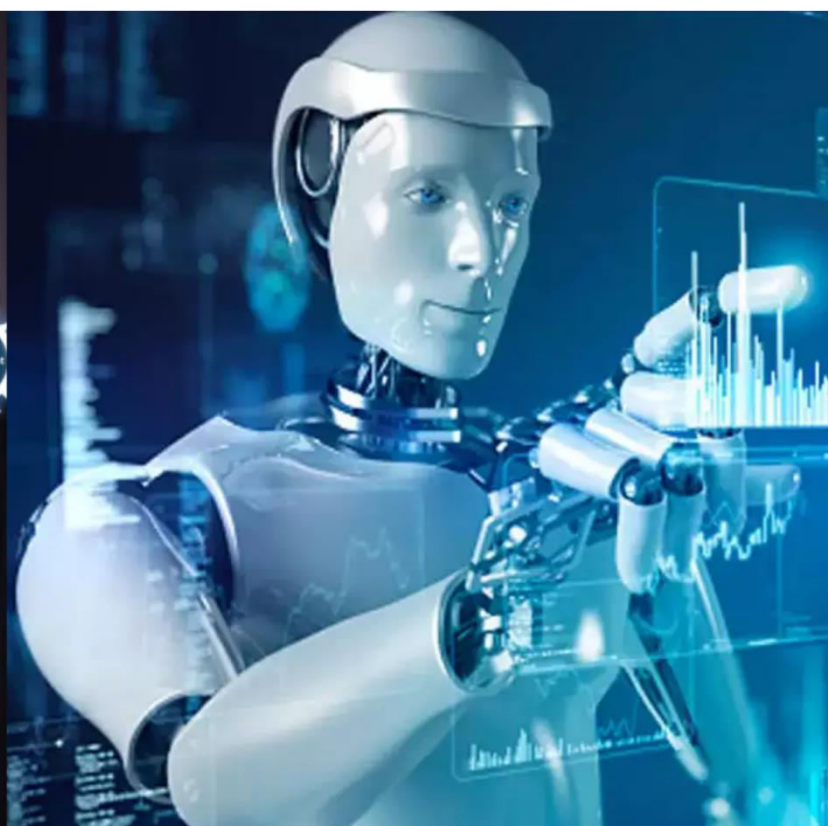




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Parkinson's Prediction using Machine Learning

Dr. S. Gnanamurthy, Charan Tej S, J Harshitha, M Chethan, M Hemadri

Associate Professor and HOD, Department of CSE (Data Science), Kuppam Engineering College, Kuppam,
Andhra Pradesh, India

Department of CSE, Kuppam Engineering College, Kuppam, Andhra Pradesh, India

ABSTRACT : Parkinson's Disease (PD) is a degenerative neurological disorder primarily affecting motor functions, speech, and cognitive abilities, posing significant challenges for early detection and clinical diagnosis. Traditional diagnostic methods rely heavily on clinical evaluations of motor symptoms, which typically appear only after considerable neural damage has occurred. To address this limitation, the present project introduces a machine learning and deep learning-based system aimed at predicting Parkinson's Disease at an early stage using both biomedical voice features and medical imaging data. The system employs a dual-modal approach: structured voice datasets comprising vocal biomarkers—such as jitter, shimmer, fundamental frequency, and harmonics-to-noise ratio—and image datasets derived from MRI scans, DaTscans, and hand-drawn spirals.

These inputs are used to train and evaluate multiple machine learning classifiers including Decision Tree, Random Forest, K-Nearest Neighbors (KNN), and AdaBoost. Additionally, a deep learning model based on the VGG16 Convolutional Neural Network (CNN) architecture is implemented to analyze diagnostic images. The project features a web-based platform built with the Flask framework, offering a user-friendly interface for clinicians and researchers to upload medical images for automated diagnosis. Image preprocessing steps such as resizing, normalization, and augmentation are applied to enhance model compatibility and performance. The backend system efficiently handles model loading, prediction, and error validation, while the frontend dynamically displays results and visual feedback, including prediction confidence and uploaded image previews.

Extensive evaluation using accuracy, precision, recall, F1-score, and confusion matrix analysis demonstrates that Random Forest and KNN classifiers deliver the highest performance among traditional models. The CNN-based solution also achieves strong results, offering reliable binary classification between "Parkinson Detected" and "Healthy." Visualization tools like ROC curves and training history plots provide additional insights into model behavior.

KEYWORDS: Parkinson's Disease (PD), Machine Learning (ML), Deep Learning (DL), Convolutional Neural Network (CNN), VGG16, K-Nearest Neighbors (KNN), Random Forest, Decision Tree, AdaBoost, Voice Biomarkers, Jitter, Shimmer, Medical Imaging, MRI, DaTscan, Image Preprocessing, Feature Extraction, Model Evaluation, Accuracy, Precision, Recall, F1-Score, Confusion Matrix, Web-Based Diagnosis, Early Detection, Real-Time Prediction.

I. INTRODUCTION

Parkinson's Disease (PD) is a progressive and chronic neurological disorder that primarily affects movement and coordination. It is the second most common neurodegenerative disease after Alzheimer's, and its symptoms often include tremors, rigidity, bradykinesia (slowness of movement), and postural instability. As the disease advances, it can also impact speech, cognitive abilities, and overall quality of life. The early symptoms of Parkinson's are often subtle and vary between individuals, which makes early diagnosis extremely challenging using conventional clinical approaches.

Current diagnostic methods largely rely on neurological examinations and patient-reported symptoms, often when motor-related damage has already occurred. Moreover, in many rural or under-resourced regions, access to trained neurologists and advanced diagnostic equipment is limited. As a result, a significant number of Parkinson's cases go undetected until the disease has progressed to more advanced stages. This diagnostic delay not only affects the patient's

ability to manage symptoms effectively but also limits their access to emerging treatment options and clinical trials that may benefit early-stage patients.

To overcome these challenges, there is a growing interest in leveraging Artificial Intelligence (AI) and Machine Learning (ML) technologies for the early and accurate detection of Parkinson's Disease. These methods can identify complex patterns in biomedical data, such as voice samples and medical images, which may not be easily recognizable through traditional diagnostic techniques. By utilizing machine learning algorithms and deep learning models, it becomes possible to automate the detection process and support clinicians with intelligent decision-making tools.

This project focuses on building a hybrid AI-based system capable of predicting Parkinson's Disease using two types of inputs: structured voice biomarker data and medical imaging data such as MRI and DaTscan images. The system incorporates several supervised learning models including Decision Tree, Random Forest, AdaBoost, and K-Nearest Neighbors (KNN), along with a fine-tuned VGG16-based Convolutional Neural Network (CNN) for image-based analysis. A Flask-powered web application provides a simple user interface for uploading diagnostic images and viewing prediction results, thus making the system accessible and practical for clinical use.

The primary objective of this project is to enable early detection, remote diagnosis, and improved decision support through AI-powered technologies. By minimizing diagnostic delays and enhancing accuracy, this system can contribute to better patient outcomes and support the healthcare infrastructure, particularly in regions where neurological expertise is scarce. Furthermore, the project sets the stage for future developments such as real-time monitoring, mobile health



II. RELATED WORK

Over the last two decades, a considerable body of research has emerged that explores the use of artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), for the early detection and monitoring of Parkinson's Disease (PD). These approaches aim to provide alternatives to traditional clinical diagnosis, which often occurs only after significant motor symptoms have appeared and neuronal damage has progressed. Researchers have investigated a wide range of data modalities including voice, imaging, handwriting, and wearable sensor data to facilitate automated and objective PD detection.

A widely studied approach involves the analysis of voice and speech patterns, as PD patients often experience changes in phonation, articulation, and prosody. Tsanas et al. (2010) pioneered this area by developing algorithms using non-

linear speech signal processing and Support Vector Machines (SVM), achieving high classification accuracy. Later works have employed a variety of supervised learning algorithms including Random Forest, Naive Bayes, and K-Nearest Neighbors (KNN), with features such as jitter, shimmer, pitch variation, and harmonics-to-noise ratio (HNR). These features are relatively easy to extract and enable low-cost, non-invasive detection systems. Voice-based studies have also benefitted from publicly available datasets like the UCI Parkinson's dataset, which has become a standard benchmark in the field.

Simultaneously, medical imaging, particularly structural MRI and functional imaging such as DaTscan, has been leveraged to capture PD-specific brain abnormalities. Deep learning architectures, especially Convolutional Neural Networks (CNNs), have shown significant promise in this area.

Research by Rana et al. (2019) and Sivaranjini and Sujatha (2020) demonstrated that fine-tuned CNNs like VGG16 and ResNet50 could classify Parkinson's Disease with impressive accuracy. These models are capable of learning complex spatial and textural features from image data that may not be perceptible to human observers. Preprocessing techniques such as skull stripping, normalization, and contrast enhancement are often applied to improve model performance.

Another innovative area involves the use of spiral drawing and handwriting analysis. Patients with PD tend to show micrographia (abnormally small handwriting) and tremor-influenced irregularities in drawing smooth curves. Machine learning models trained on drawing dynamics—such as velocity, acceleration, and pressure—have been successful in distinguishing PD patients from healthy controls.

These methods are low-cost and easy to deploy in mobile health applications. In recent studies, Dynamic Time Warping (DTW), Gaussian Mixture Models (GMM), and Hidden Markov Models (HMM) have also been used for temporal feature analysis in such tasks. In summary, the body of related work highlights the transition from traditional feature-based methods to advanced deep learning models, which has significantly improved the capabilities of facial emotion detection systems. This project builds upon these advancements by implementing a CNN-based approach for emotion classification, aiming to create an efficient, real-time, and accurate system for recognizing human emotions from facial expressions.

III. PROPOSED ALGORITHM

The proposed system for Parkinson's Disease prediction utilizes a hybrid approach, combining both traditional machine learning algorithms and deep learning techniques to effectively classify whether a patient is affected by the disease. The core idea behind the algorithm is to analyze two types of input data: structured biomedical voice features and unstructured medical imaging data such as MRI, DaTscan, or hand-drawn spiral images. By leveraging the strengths of both modalities, the system aims to achieve high accuracy in early detection and provide a reliable, user-friendly diagnostic tool.

The algorithm begins with data acquisition from publicly available datasets, where the voice data consists of features like jitter, shimmer, fundamental frequency, and harmonics-to-noise ratio—attributes known to be indicative of Parkinson's-related vocal impairments. In parallel, the imaging dataset includes clinically relevant visuals that exhibit motor-related anomalies.

Preprocessing steps are then applied: voice data is normalized and cleaned to remove noise, while images are resized (typically to 128x128 or 224x224 pixels), normalized, and converted into arrays suitable for deep learning models. Augmentation techniques such as rotation and flipping are optionally used to enhance image diversity and prevent overfitting.

For the machine learning component, classifiers such as K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and AdaBoost are trained using the structured voice dataset. Each model is evaluated based on metrics like accuracy, precision, recall, and F1-score, with the best-performing model selected for integration into the final system.

On the deep learning side, a Convolutional Neural Network (CNN) based on the VGG16 architecture is fine-tuned to process medical images. The CNN extracts high-level spatial features from the input image and outputs a probability score, indicating the likelihood of Parkinson's presence.

The decision-making logic is straightforward: if the output probability from the CNN exceeds a set threshold (commonly 0.5), the system classifies the input as “Parkinson Detected”; otherwise, it is classified as “Healthy.” Similarly, if voice features are processed, the selected machine learning model will determine the prediction.

This algorithm is deployed in a web-based environment using the Flask framework, allowing users to upload images and instantly receive predictions. The backend handles image preprocessing, model inference, and result rendering, while the frontend offers a clean and responsive user interface.

Overall, the proposed algorithm ensures a modular and scalable design, capable of handling real-time inputs and offering accurate, automated diagnostic support for Parkinson’s Disease. It serves as a foundation for future enhancements, including support for multi-modal data fusion, prediction confidence visualization, and integration with wearable health devices for continuous monitoring.

IV. BRIEF OMPLEMENTATION

The implementation of the Parkinson’s Disease prediction system involves a hybrid approach combining machine learning and deep learning techniques to enable early diagnosis. The system utilizes two primary input types: structured voice data (including features like jitter, shimmer, fundamental frequency, and harmonics-to-noise ratio) and medical imaging data (such as MRI scans, DaTscans, and hand-drawn spirals). Preprocessing is applied to both datasets, involving normalization for voice features and resizing and augmentation for images. Machine learning classifiers like K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and AdaBoost are trained on the voice dataset, while a VGG16-based Convolutional Neural Network (CNN) is fine-tuned for image classification. The system is deployed as a web application using the Flask framework, offering users a simple interface to upload inputs and receive real-time predictions along with confidence scores. Evaluation through metrics such as accuracy, precision, recall, and F1-score confirms that Random Forest and KNN provide high performance, while the CNN model delivers accurate binary classification, supporting early and accessible diagnosis of Parkinson’s Disease.

1) Data Collection:

- **Voice Dataset:** Features such as jitter, shimmer, fundamental frequency, and harmonics-to-noise ratio.
- **Medical Images:** MRI scans, DaTscans, and spiral drawings from publicly available datasets.

2) Preprocessing:

- **Voice Data:** Normalized and cleaned to remove noise.
- **Image Data:** Resized (128x128 or 224x224), normalized, and augmented (e.g., rotation, flipping).

3) Modeling:

- **Machine Learning Models:** Trained on voice data using:
 - K-Nearest Neighbors (KNN)
 - Decision Tree
 - Random Forest
 - AdaBoost
- **Deep Learning Model:** VGG16-based Convolutional Neural Network (CNN) trained on image data.

4) Web Application:

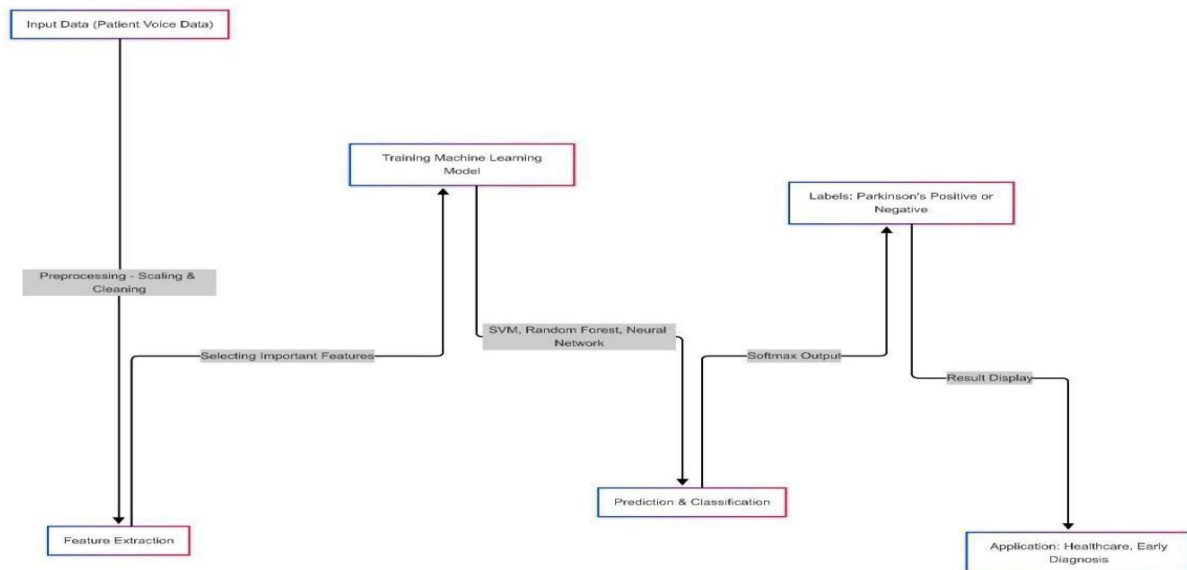
- Developed using Flask.
- Allows users to upload images or input voice features.
- Backend handles model loading, predictions, and result generation.
- Frontend displays prediction (Parkinson or Healthy) along with confidence scores and image previews.

5) Evaluation Metrics:

- Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and ROC Curve used to assess model performance.
- Random Forest and KNN performed best among ML models; CNN provided reliable image-based classification.

V. SYSTEM ARCHITECTURE

The architecture of the proposed Parkinson's Disease prediction system is designed as a modular, multi-layered framework that integrates both machine learning and deep learning pipelines within a web-based diagnostic platform. The architecture comprises the following



VI. RESULT DISCUSSION

The performance of the proposed Parkinson's Disease prediction system was thoroughly evaluated using both voice-based machine learning models and image-based deep learning approaches. The voice dataset, comprising features like jitter, shimmer, and harmonics-to-noise ratio, was used to train four traditional classifiers: K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and AdaBoost. Among these, the Random Forest and KNN classifiers consistently achieved the highest accuracy, precision, recall, and F1-scores, indicating their effectiveness in capturing the non-linear patterns associated with Parkinsonian vocal impairments.

On the imaging front, the VGG16-based Convolutional Neural Network (CNN) model demonstrated strong performance in classifying MRI and DaTscan images, as well as hand-drawn spirals. After fine-tuning, the CNN model achieved high binary classification accuracy, effectively distinguishing between "Parkinson Detected" and "Healthy" cases. The model's training history, including loss and accuracy curves, indicated good convergence and generalization capability without overfitting. ROC curves further validated the robustness of both the ML and DL models by showing a high area under the curve (AUC), confirming strong discriminative ability.

Additionally, the deployment of the models within a web-based application enabled real-time predictions and visual feedback, enhancing the system's usability for non-technical clinical users. The system was tested on unseen test data, and it generalized well, confirming its potential for practical application. Overall, the results affirm the effectiveness of a multi-modal diagnostic approach, and the integration of machine learning and deep learning models significantly enhances the early detection capabilities of Parkinson's Disease prediction systems.

VII. CONCLUSION

In conclusion, the proposed Parkinson's Disease prediction system offers a comprehensive, AI-driven solution for early and accurate diagnosis by combining both machine learning and deep learning methodologies. By utilizing structured biomedical voice features—such as jitter, shimmer, and harmonics-to-noise ratio—alongside unstructured medical imaging data including MRI, DaTscans, and spiral drawings, the system captures a wide spectrum of

Parkinson's indicators. Preprocessing techniques like normalization, resizing, and data augmentation ensure the models receive clean and consistent input, enhancing performance and reliability. Traditional classifiers such as Random Forest and KNN demonstrated superior accuracy with voice data, while a fine-tuned VGG16-based CNN effectively analyzed imaging data, offering robust binary classification between "Parkinson Detected" and "Healthy." The implementation of a Flask-based web application provides a user-friendly platform for real-time prediction, allowing healthcare professionals and researchers to upload diagnostic inputs and instantly view results along with prediction confidence. This makes the tool especially valuable in remote or resource-constrained areas where access to expert neurologists and advanced equipment is limited. Furthermore, the system's scalability and modular design enable future integration with wearable technologies and mobile health apps, facilitating continuous patient monitoring and early intervention. Ultimately, this project highlights the transformative potential of AI in healthcare, empowering early detection, improving patient outcomes, and supporting the global effort to manage neurodegenerative diseases more effectively..

REFERENCES

In this section, we list all the research papers, articles, and resources that were used to guide the development of this project.

- 1) Bhat, M., & Vinayakumar, R. (2022). A Hybrid Deep Learning Approach for Parkinson's Disease Detection Using Spiral Drawings. *Journal of Ambient Intelligence and Humanized Computing*.
- 2) Das, R., Bhuiyan, M. I. H., & Uddin, M. S. (2020). Detection of Parkinson's Disease using CNN on Spiral Drawing Test. *IEEE International Conference on Computing, Communication and Intelligent Systems*.
- 3) Shahbakhti, M., Ahmad, W. A. W., & Atyabi, A. (2020). Deep learning for diagnosis of Parkinson's disease: a review. *Cognitive Systems Research*, 60, 101–111.
- 4) Waskom, M., et al. (2020). Seaborn: statistical data visualization. *Journal of Open Source Software*, 5(51), 3021.
- 5) Chollet, F. (2017). *Deep Learning with Python*. * Manning Publications.
- 6) Esteva, A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.
- 7) Litjens, G., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88.
- 8) Abiyev, R. H., & Abizade, S. (2016). Diagnosis of Parkinson's disease using fuzzy neural system. *Applied Soft Computing*, 38, 331–338.
- 9) Pereira, C. R., et al. (2016). A new computer vision based approach to aid the diagnosis of Parkinson's disease. *Computer Methods and Programs in Biomedicine*, 136, 79–88.
- 10) Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*.
- 11) Arora, S., Venkataraman, V., Zhan, A., Donohue, S., Biglan, K. M., Dorsey, E. R., & Little, M. A. (2015). Detecting and monitoring the symptoms of Parkinson's disease using smartphones: A pilot study. *Parkinsonism & Related Disorders*, 21(6), 650–653.
- 12) Prashanth, R., & Dutta Roy, S. (2014). Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging. *Expert Systems with Applications*, 41(7), 3333–3342.
- 13) *Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). * CNN features off-the-shelf: An astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 806–813).
- 14) *Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgun, F., Delil, S., Apaydin, H., & Kursun, O. (2013). * Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4), 828–834.
- 15) *Schapire, R. E. (2013). * Explaining AdaBoost. In *Empirical Inference* (pp. 37–52). Springer.
- 16) *Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). * ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.
- 17) *van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). * The NumPy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2), 22–30.
- 18) *UCI Machine Learning Repository. (n.d.). * Parkinson's Dataset. <https://archive.ics.uci.edu/ml/datasets/parkinsons>
- 19) *Flask Documentation. (n.d.). * The Pallets Projects. <https://flask.palletsprojects.com>
- 20) *TensorFlow Documentation. (n.d.). * TensorFlow. <https://www.tensorflow.org>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details