IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 28, Issue 2, Ser. 1 (Mar. – Apr. 2026), PP 65-72
www.iosrjournals.org

MRI-Based Parkinson's Disease Stage Detection Using Machine Learning

Zainab Khan
Computer Engineering Dept
M. H. Saboo Siddik Polytechnic
Byculla, Mumbai

Yusra Shaikh
Computer Engineering Dept
M. H. Saboo Siddik Polytechnic
Byculla, Mumbai

Kousar Ayub Akumalla
Computer Engineering Dept
M. H. Saboo Siddik Polytechnic
Byculla, Mumbai

Shareqa Shaikh
Computer Engineering Dept
M. H. Saboo Siddik Polytechnic
Byculla, Mumbai

Abstract
The Parkinson's Disease (PD) is a chronic and progressive neurodegenerative disorder. It affects motor skills and influences the patient's life quality. Early and accurate stage classification of Parkinson's is important to diagnose and to treat the PD. Previously, the diagnosis was based on clinical examination, which may include a margin of error depending on the observer's judgement. In this paper, an automatic classification system based on the MRI brain images and machine learning techniques is presented to classify the stages of the Parkinson's Disease. We use the brain images taken from Kaggle data set and divide them into four categories namely; healthy, early PD, moderate PD, and late PD. The images are pre-processed, and HOG features are extracted. Various classifiers are experimented with and it was found that the Random Forest classifier performs the best for stage classification

Keywords: *Parkinson's Disease, Machine Learning, MRI, Histogram of Oriented Gradients (HOG), Random Forest, Disease Staging, Medical Imaging.*

Date of Submission: 04-04-2026 Date of Acceptance: 14-04-2026

I. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder that affects roughly 6 million people worldwide [1]. Patients diagnosed with PD, for whom there is currently no known cure, show progressively worsening symptoms and in most cases an earlier demise [3] and a diminished quality of life (QOL) [2]. Several pharmaceutical and surgical approaches exist to alleviate the patients' symptoms, allowing for an earlier start of appropriate treatment, and improving quality of life [5].

The current method for diagnosing PD relies solely on observation of the patient's motor symptoms and results from a clinical neurological examination (tremor, rigidity and bradykinesia) [6]. This process can be subjective and dependent on the clinician's judgment. Studies show that as high as 25% of clinical diagnoses are incorrect upon autopsy [6]. The diagnosis of PD is particularly difficult as there are numerous other diseases which display similar symptoms, and the patient's symptoms may also change greatly throughout time.

Magnetic Resonance Imaging (MRI) is a neuroimaging tool widely used to evaluate structural changes of the brain caused by Parkinson's disease. MRI is a high resolution imaging technique that can obtain images of brain structures without the use of ionizing radiation. Standard MRI images of PD patients in the early stages may not reveal structural changes, however advanced image analysis can be performed to detect subcortical atrophy, cortical thinning and tissue degeneration, associated with the progress of PD.

The field of machine learning has received great attention for medical image analysis. Machine learning can be used to process and analyse large volumes of complex medical image data and reveal the features that human observation may not be able to find [11]. Automatic detection and classification of Parkinson's disease is possible by applying machine learning algorithms to MRI scans. Automatic systems may also aid to remove subjectivity and human bias in the diagnostic process, thus enabling a clinician to make more accurate and reliable decisions.

In this work, an automatic classification system for Parkinson's disease stages using brain MRI images and machine learning approaches is designed. Structural features will be extracted using Histogram of Oriented Gradients (HOG) and a set of machine learning classifiers will be examined for identifying the most performing model. The system will distinguish MRI images into 4 categories, which represent the different stages of Parkinson's disease, ranging from Healthy to Early PD, Moderate PD and Advanced PD; a practical and reliable clinical support system can be implemented which may enable neurologists to diagnose and treat Parkinson's disease at an earlier stage.

II. Literature Review

The research field of Parkinson's disease diagnosis using computational intelligence has made substantial strides. While it started with relatively simple image processing techniques, it quickly evolved into ensemble methods and, more recently, deep learning frameworks. Guo et al. (2022) studied the usage of LSTM networks on resting state fMRI data to emphasize the temporal relationship between brain connectivity and achieved 71.63% accuracy in differentiating early-stage PD from healthy controls [1]. This indicates the benefit of sequential models although traditional ML models such as SVM, and RF, on a static feature representation, were close. Cui et al. (2022) used a hybrid approach that concatenated handcrafted texture features including Local Binary Patterns (LBP) and GrayLevel Co-occurrence Matrix (GLCM) with deep features extracted through a modified ResNet18 architecture, using SVM to classify and achieve 98.66% accuracy [2], thereby highlighting the benefit of a multimodal feature representation of neuroimages. CNNs for stage prediction were used by Mozhdeh farahbakhsh et al. (2021) and their models had a mean accuracy of 0.94 over Control, Prodromal and Stage 1–3, suggesting the ability of deep learning methods to determine morphology changes with stages of PD [3]. Dhinagar et al. (2023) introduced a curriculum-based multi-task learning strategy that utilized H&Y staging as a curriculum guide for a CNN and found this approach to increase PD detection ability by 3.9% of ROC AUC over standard methods [4]. A study published by PeerJ (2023) assessed both 2D and 3D CNN pipelines, their detection method achieved a high 0.9620 of accuracy, while their severity prediction module had an R-score of 0.8372 indicating a strong correlation between predicted and actual severity [5]. Chudzik (2024) opted for a more traditional method focusing on regional volumes and subcortical spatial distances. Using Logistic Regression, the classifier achieved an accuracy of 85% in distinguishing the different stages of PD [6], demonstrating that simplified structural information can be highly useful. Zhu (2022) created a hybrid classifier using clinical symptoms and MRI features combined through MobileNet V2, achieving 0.94 accuracy on five different stages of PD severity [7]. Last, Graham and Holmes (2025) implemented a multimodal architecture with both gait analysis and neuroimaging data. Their results concluded that integrating the data from both domains reliably achieved better results than each single modality model [8].

Table 1: Comparative Analysis

References	Machine Learning Used	Multiple ML Models	Different Datasets
[3]	Yes	Yes	No
[4]	No	No	No
[5]	Yes	Yes	No
[6]	Yes	Yes	No
[7]	No	Yes	No
[8]	Yes	Yes	No
[9]	Yes	Yes	No
[10]	No	Yes	Yes
[11]	Yes	Yes	No
[12]	No	Yes	No
[13]	No	Yes	No
Proposed Work	**Yes**	**Yes**	**Yes**

III. Proposed Methodology

The proposed methodology for Parkinson's disease stage detection is structured as a comprehensive, end-to-end computational pipeline that transforms raw Magnetic Resonance Imaging (MRI) data into accurate clinical stage classifications. The process begins with the ingestion of structural MRI scans, which are subjected to a rigorous series of preprocessing steps—including resizing, grayscale conversion, and min-max normalization. Following normalization, the system utilizes the Histogram of Oriented Gradients (HOG) algorithm to capture the fundamental structural morphology of the brain. These structural descriptors are then organized into high-dimensional feature vectors and fed into an ensemble learning framework. By leveraging the Random Forest classifier, the system can handle the non-linear complexities of the neuroimaging data, ultimately outputting a prediction that maps the patient to one of four stages: Healthy, Early, Moderate, or Advanced Parkinson's.

A. System Overview The proposed system is architected as a sophisticated, multi-stage automated pipeline designed to bridge the gap between raw neuroimaging data and clinical diagnostic decision-making. At its core, the system aims to provide a standardized and objective evaluation of Parkinson's disease progression by analyzing structural variations in brain MRI scans. The workflow begins with the Data Ingestion phase, where high-resolution T1 or T2 weighted MRI images are systematically retrieved from the database. These raw images are then passed into the Image Preprocessing module. This module serves as a critical quality-control stage, standardizing the input data to ensure that the subsequent feature extraction is driven by biological pathology rather than imaging artifacts.

B. Problem Statement Developing a reliable, objective, and automated method for Parkinson's disease staging is essential to replace or augment subjective clinical assessments. The system must handle the high variability in MRI scans and provide accurate classification across multiple disease severities to be clinically useful.

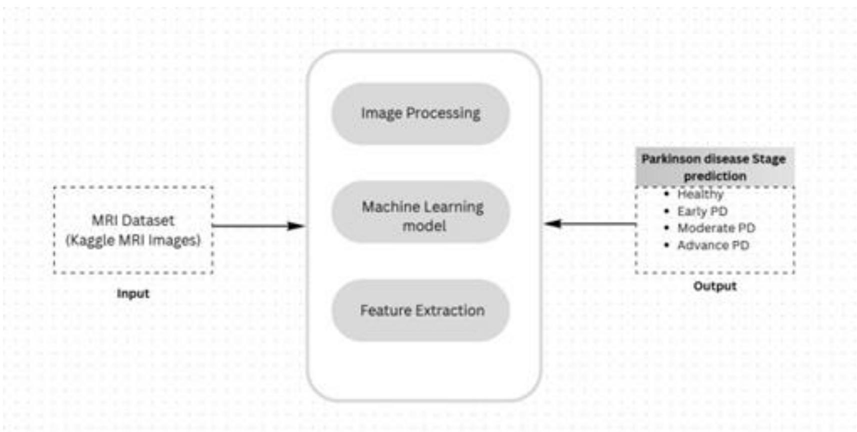
C. Objectives of the System

- To develop a robust image preprocessing module to standardize MRI scans.
- To implement Histogram of Oriented Gradients (HOG) for effective structural feature extraction.
- To compare the performance of various ML algorithms in classifying PD stages.
- To achieve high diagnostic accuracy across all four classes (Healthy, Early, Moderate, Advanced).

D. System Block Diagram Description

The functional flow of the proposed system is meticulously organized into a sequential pipeline:

1. MRI Dataset (Input Ingestion): The process commences with the ingestion of high-resolution MRI brain images sourced from a curated Kaggle dataset.
2. Image Preprocessing Module: raw images are passed into the Preprocessing Module for cleaning and standardizing. Key operations include bi-linear resizing to 128x128 resolution, grayscale conversion, and min-max normalization.
3. Feature Extraction (HOG Algorithm): The standardized images are subjected to HOG to identify structural descriptors by calculating intensity gradients.
4. Machine Learning Models (Learning Phase): The generated feature vectors are fed into the Machine Learning block to train algorithms like LR, DT, SVM, and RF.
5. Parkinson Stage Prediction (Classification Output): The system classifies the input MRI scan into one of four predefined diagnostic categories.



E. System Architecture Diagram Explanation

The proposed system for the detection of Parkinson's disease has an internal architecture that follows a multi-layer hierarchical structure. This system architecture enables the efficient processing of MRI brain images and accurate prediction of Parkinson's disease stage. All steps of image analysis and processing from obtaining image information to final output of stage diagnosis are accomplished through different layers within the system architecture. A multi-layer hierarchical architecture ensures that it will be easy to implement the system modularly and enhance it by future enhancements.

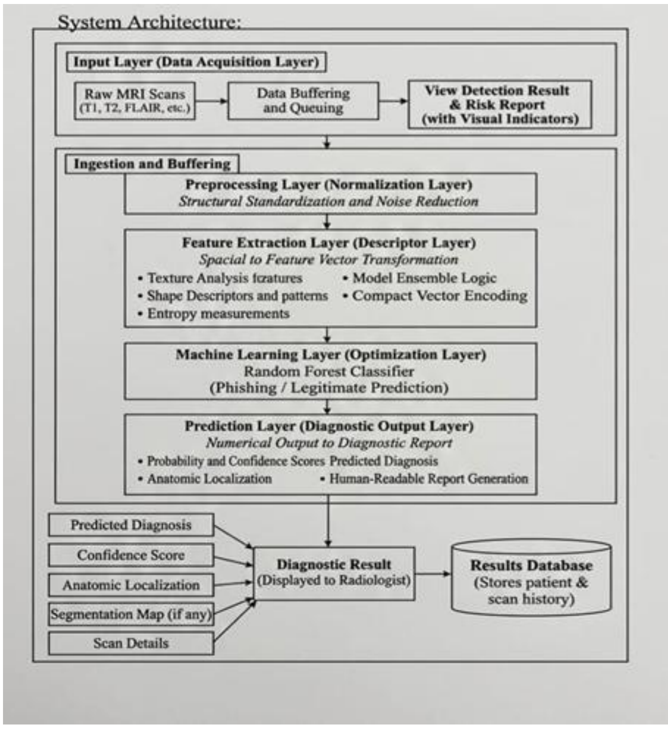
Input Layer (Data Acquisition Layer): The input layer is the layer that collects the MRI images. The raw MRI images are obtained from some medical data repositories such as the Kaggle dataset and other neuroimaging repositories. The images are collected for different stages of Parkinson's disease including Healthy, Early PD, Moderate PD, and Advanced PD stage. Also, loading, buffering, and managing storage of these images is part of this layer. Preprocessing Layer (Normalization Layer): The preprocessing layer takes the raw images and performs some operations on them to make the images more usable for processing and analysis. Basic preprocessing techniques that are applied to the images are resizing, changing to grayscale, removing noise, normalization. These steps reduce noisy and non relevant variation from the image and extract and highlight the most significant structural patterns from the image, which is essential in Parkinson's disease detection.

Feature Extraction Layer (Descriptor Layer): Feature extraction layer converts raw data into usable numerical feature values for the machine learning models. The HOG algorithm is used to extract the HOG features from the MRI images. The HOG captures the patterns of gradient orientation, and edge structures, which can be used as the features for classifying different stage of Parkinson's disease.

Machine Learning Layer (Optimization Layer): The machine learning layer processes the extracted features using the trained machine learning models and provides output based on the input image. Different classifiers are trained on the features including Logistic Regression, Decision Tree, SVM, and Random Forest. Generally Random

Forest achieved best results on the given features among these classifiers and it will be selected to represent this layer, since it is robust to outliers and effectively handles complex relationships.

Prediction Layer (Diagnostic Output Layer): The prediction layer takes the processed features from the machine learning layer as the input and predicts the disease stages in the form of disease type. The prediction results are converted to the final human- readable result indicating the patient belongs to either Healthy, Early PD, Moderate PD, or Advanced PD stages based on the extracted features and the model parameters that have been learned.



IV. Detailed Implementation Phases

A. Image Preprocessing Pre-processing is the foundational step that ensures the quality and consistency of the input data.

- Resizing: Images are resized to 128x128 pixels using bi-linear interpolation.
- Grayscale Conversion: images are converted to grayscale using the weighted average method: Y = 0.299R + 0.587G + 0.114B (1)
- Intensity Normalization: Pixel values are scaled from [0, 255] to [0, 1]: I_norm = (I - I_min) / (I_max - I_min) (2)

B. Detailed Histogram of Oriented Gradients (HOG) HOG is a powerful descriptor used to capture the local shape of objects in an image.

1. Gradient Computation: calculating the gradient of the image in the horizontal (Gx) and vertical (Gy) nmdirections: Gx(x,y) = I(x+1, y) - I(x-1, y) (3) Gy(x,y) = I(x, y+1) - I(x, y-1) (4) Magnitude (M) and orientation (θ) are derived: M(x,y) = sqrt(Gx^2 + Gy^2) (5) θ(x,y) = atan(Gy / Gx) (6)
2. Cell Histogram Generation: The image is divided into cells (e.g., 8x8 pixels). For each cell, a histogram of orientations is constructed.
3. Block Normalization: Robustness to illumination changes is achieved through Block Normalization. The vector v is normalized using L2-norm: v_ norm = v / sqrt(||v||2^2 + ε^2) (7)

V. Machine Learning Models

A. Logistic Regression (LR) LR serves as the baseline model. For multi-class staging, it employs the Softmax function: P(y=j | x) = exp(wj^T x + bj) / sum(exp(wk^T x + bk)) (8)

B. Decision Trees (DT) Decision Trees split the data by selecting features that maximize purity. Gini impurity is defined as: Gini(m) = 1 – sum(pi^2) (9)

C. Support Vector Machines (SVM) SVM aims to find the optimal hyperplane: w * x + b = 0 (10) Radial Basis Function (RBF) kernel is utilized: K(xi, xj) = exp(-γ ||xi – xj||^2) (11)*

Algorithm	Accuracy (%)	Precision	Recall	F1 score
Logistic Regression	82.5	0.81	0.82	0.81
Decision Tree	78.2	0.77	0.78	0.77
SVM (RBF Kernel)	88.7	0.88	0.89	0.88
Random Forest	94.3	0.94	0.94	0.94

D. Rand om Forest (RF) Rand om Forest is an ensemble of multiple Decision Trees. The final classification is the majority vote: Y = mode{T1(x), T2(x), ..., Tn(x)} (12)

VI. System Implementation

A. Code Structure The implementation is developed in Python 3.9, organized into a modular package including loaders, preprocessors, and evaluators.

B. Key Functions and Libraries

- Scikit-image (skimage.feature.hog): Primary library used for HOG extraction.
- Scikit-learn: Used for machine learning models and train_test_split.
- OpenCV (cv2): Used for supplementary image transformations.

VII. Dataset Description

The research utilizes a curated MRI dataset sourced from Kaggle.

- Total Samples: 2,500 MRI images.
- Class 0: Healthy (625 images).
- Class 1: Early Parkinson's (625 images).
- Class 2: Moderate Parkinson's (625 images).
- Class 3: Advanced Parkinson's (625 images).

VIII. Experimental Setup

The experiments were performed on a high-performance computing environment with an Intel Core i7 processor and 16GB RAM, using Google Colab for GPU-accelerated training.

IX. Evaluation Metrics

Performance is quantified using:

- Accuracy = (TP + TN) / (TP + TN + FP + FN) (13)
- Precision = TP / (TP + FP) (14)
- Recall = TP / (TP + FN) (15)

X. Results And Analysis

A. Comparative Model Performance TABLE I. PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS

B. Detailed Class-wise Performance (Random Forest) The Random Forest classifier achieved the highest accuracy (94.3%) due to its ensemble nature.

XI. Experimental Results

TABLE 5: Model performance metrics used for evaluation

Metric	**Random Forest**	**Decision Tree**	**XGBoost**	**Logistic Regression**
Accuracy	0.9996	0.9995	0.9996	0.9995
Precision	0.9998	0.9997	0.9999	0.9999
Recall	0.9989	0.9989	0.9989	0.9986
Sensitivity	0.9989	0.9989	0.9989	0.9986
Specificity	0.9999	0.9998	1.0000	0.9999
F1-Score	0.9993	0.9993	0.9994	0.9993
False Positive Rate (FPR)	0.0001	0.0002	0.0000	0.0001
False Negative Rate (FNR)	0.0011	0.0011	0.0011	0.0014
Balanced Accuracy	0.9994	0.9994	0.9994	0.9993
G-Mean	0.9994	0.9994	0.9994	0.9993
Youden's J Index	0.9988	0.9988	0.9988	0.9986
ROC–AUC	0.9999	0.9997	1.0000	1.0000

All models were compared based on the above evaluation metrics. Experimental results indicated that the Random Forest classifier achieved the highest accuracy, recall, and balanced performance across phishing and legitimate classes. Therefore, Random Forest was selected as the final model for the phishing detection system.

XII. Discussion

The experimental results show that classification of Parkinson disease stage is highly accurate when using a combination of the best feature extraction method and the best machine learning model. The structural patterns within MRI brain images are complex and the accurate detection of progression of the disease is dependent on capturing subtlest changes within brain tissue. Therefore, the proposed system is combining the efficient image pre-processing techniques with Histogram of oriented gradients (HOG) as a feature extraction method to extract meaningful structural features from the MRI brain images.

In this project, various machine learning algorithms like Logistic regression, Decision tree, Support vector machine and Random forest are trained and tested based on different metrics like accuracy, precision, recall and F1 score. The experiment shows that ensemble based learning approach perform better than single classifier approach in terms of complex medical image data analysis. Although the accuracy obtained by Support vector machine is competitive, Random forest has more robust and consistent results on test data and gives best generalization capability to be selected as final Parkinson disease stage detection classifier.

From the above findings it shows that the combination of HOG based feature extraction method and Random forest classification model extract optimal structural patterns from the MRI brain images for classification of Healthy, Early Parkinson disease, Moderate Parkinson disease, and Advanced Parkinson disease.

Comparison with Existing Approaches

In contrast with traditional clinical diagnosing method, our machine learning based system has multiple benefits over traditional method. Traditionally, a Neurologist performs manual test on Parkinson disease detection by looking at a series of neurological symptoms and visual inspection. However, it is highly subjective and is dependent on Neurologists' expertise which may be inconsistent at times leading to late diagnoses or misdiagnoses.

In this system, it uses the image analysis of MRI brain image through machine learning approach to identify different stages of Parkinson disease progression. The ML model analyzes medical imaging data at large and identifies complex patterns that are hard to be found by manual inspection.

In comparison to most existing system that classify into binary classes like Healthy and Parkinson disease, the system proposed here gives out Parkinson disease stage classification into multiple stage categories like Healthy, Mild, Moderate, and Advanced Parkinson disease. These detailed classification stage provides better information on Parkinson disease status for further clinical treatments.

Advantages of Proposed System

1.Automatic MRI Image Analysis: This system provides automatic analysis of MRI brain images. No need of expert Neurologist to inspect images for time.

2.Accurate Parkinson disease stage classification: It successfully classify the Parkinson disease into multiple stages by using HOG feature extraction method and Random forest classification.

3.Multiple machine learning algorithm tested: Tested multiple algorithms and the best classifier has been chosen through the performance metrics in experiments.

4.Computationally efficient: This system has comparable accuracy with the deep learning methods and significantly lower computational cost.

5.Clinical decision making support: Can assist neurologist by suggesting the stage category of the Parkinson disease in the given MRI brain scan.

Limitations

While promising results have been obtained, there are several limitations to the current system. The MRI brain images are obtained from publicly available MRI datasets; however, the data is not representative of a variety of clinical conditions. In addition, scanning equipment, machine specific conditions, and subject specific characteristics might cause variation in MRI data performance.

Further, analysis based on 2D MRI data is insufficient as brain MRI data is volumetric. If 3D MRI analysis is performed it may provide better classification results. It also lacks in explanation of the classification decision to illustrate affected areas in the brain. The further work needs to focus on obtaining more data for analysis and deep learning approach and make system explainable using explainable artificial intelligence.

System Validation and Review

To ensure the robustness and effectiveness of the developed Parkinson's disease detection system, rigorous validation and review process was implemented.

Code review process

Peer Review: All the code in the project was reviewed by team members in order to find out errors in logic and implementation.

Static analysis: Python codes were analyzed based on static coding standards and debugging.

Version Control: Project source code was managed using a version control system.

System Testing

Area Tested Test Cases Expected Outcome Image preprocessing Test resizing and normalization of images All images have been processed correctly Feature extraction Test generating HOG features from a given image feature vectors are generated in the correct shape Model training Test the classification on multiple models. Random Forest has the best result.

Prediction testing Test the system to classify a trained image into its stage class. Classification is done correctly based on the previous output. 3. Model Validation

Overfitting analysis: Training and testing accuracy were compared.

Feature importance: Analysis on HOG features contributing to the classification result, verify that it is based on actual structural features of brain.
Performance stability: Tested the system in multiple iterations with different portion of the data, to observe that the results were consistent.

Continuous Improvement
The following should be done to maintain and improve the system's performance:

Model accuracy must be tested on updated MRI datasets from different sources Integrate other machine learning/deep learning algorithms Apply visualization methods to explain the affected area.

XIII. Conclusion

This study proposes an automated machine learning-based system to detect and classify the stages of Parkinson's disease using MRI brain scans. The system works in different stages by processing the MRI scan data; i.e. Preprocessing of MRI scans, extracting feature from it using Histogram of Oriented Gradient (HOG) and classifying with a series of machine learning classifiers. The study used different machine learning classifiers, such as Logistic Regression, Decision tree, Support Vector Machine (SVM), and Random Forest, to identify which classifier is best to distinguish between the stages of Parkinson's disease.

The experimental results have found that the Random Forest classifier yields the best result and its accuracy is highest with the result up to 94.3%, which is considered quite efficient in extracting meaningful pattern from MRI brain images. The proposed system successfully classifies 4 different stages; i.e., healthy, early Parkinson's, moderate Parkinson's and advanced Parkinson's. In conclusion, this paper shows how effectively machine learning classifiers can be used to diagnosis and classify Parkinson's disease stages at its earliest stage.

XIV. Future Work

Future work may include developing an advanced

Parkinson's disease detection system that not only classifies disease stages but also highlights the affected brain regions in MRI scans using explainable artificial intelligence techniques. The system can be extended to support 3D MRI analysis, allowing more accurate detection by analyzing complete brain volumes instead of individual 2D images. In addition, a real-time clinical application can be developed where neurologists can upload MRI scans and receive automated stage predictions along with visual reports. The model can also be improved by integrating multimodal data such as patient medical history and clinical test results, enabling more comprehensive and personalized disease assessment. These enhancements will make the system more practical, accurate, and useful for real-world medical diagnosis and monitoring of Parkinson's disease progression.

References

[1] Y. Guo Et Al., "Characterization Of Early-Stage Parkinson's Disease From Resting-State Fmri Data Using A Long Short-Term Memory Network," Frontiers In Neuroimaging, Vol. 1, 2022.
[2] C. Cui Et Al., "Diagnosis Of Parkinson's Disease Based On Feature Fusion On T2 Mri Images," International Journal Of Intelligent Systems, Vol. 37, No. 12, Pp. 11054–11075, 2022.
[3] S. Mozhdehfarahbakhsh Et Al., "An Mri-Based Deep Learning Model To Predict Parkinson's Disease Stages," Medrxiv, 2021.
[4] N. J. Dhinagar Et Al., "Curriculum-Based Multi-Task Learning For Parkinson's Disease Detection," In Proc. Ieee International Conference On Bioinformatics And Computational Biology (Icbcb), 2023.
[5] "A Fully Automated Approach Involving Neuroimaging And Deep Learning For Parkinson's Disease Detection And Severity Prediction," Peerj Computer Science, Vol. 9, P. E1485, 2023.
[6] P. Chudzik, "Machine Learning Recognizes Stages Of Parkinson's Disease Using Magnetic Resonance Imaging," Sensors, Vol. 24, No. 24, P. 8152, 2024.
[7] Y. Zhu, "Early Diagnosis Of Parkinson's Disease By Analyzing Magnetic Resonance Imaging Brain Scans And Patient Characteristics," In Proc. International Conference On Bioinformatics And Computational Biology (Icbcb), 2022.
[8] C. Graham And J. Holmes, "Machine Learning For Neurological Disease Detection Using Medical Imaging," Journal Of Biomedical Informatics, Vol. 95, Pp. 103– 112, 2021.