



SASTRA Deemed to be University  
Knowledge Connect @ SASTRA

---

Theses and Dissertations

---

17-7-2025

## A Predictive Framework for Early Detection and Personalised Monitoring of Parkinson's Disease Using Artificial Intelligence and Large Language Models

Priyadarshini S Ms

Follow this and additional works at: <https://knowledgeconnect.sastra.edu/theses>

 Part of the [Other Computer Engineering Commons](#)

---

**A Predictive Framework for Early Detection and Personalised  
Monitoring of Parkinson's Disease Using Artificial Intelligence and  
Large Language Models**

A Thesis

Submitted to the Faculty

of

SAASTRA Deemed University

by

Priya Dharshini S

in partial fulfilment of the

requirements for the degree

of

Doctor of Philosophy

August 2025

## **DEDICATIONS**

I dedicate this thesis to my beloved family & friends

## ACKNOWLEDGMENTS

First, I express my gratitude to the Almighty for bestowing me with the strength, knowledge, and opportunity to take this research journey and see it through to a satisfactory completion. Beyond my efforts, the success of this thesis hinges significantly on the support and guidance of numerous individuals.

I deeply thank **Prof. R. Sethuraman**, Chancellor, SASTRA University, for his unwavering support and for having provided me with excellent facilities to carry out my research. I wish to express my gratitude to **Prof. S. Vaidhyasubramaniam**, Vice-Chancellor, **Dr. R. Chandramouli**, Registrar, **Dr. S. Swaminathan**, Director-CeNTAB, and Dean (Planning and Development), **Dr. John Bosco Balaguru**, Dean (Sponsored Research), **Dr. K. Sridhar**, Dean (Research), **Dr. Adline Princy S**, (Associate Dean-Research) **Dr. K. Thenmozhi**, Dean (SEEE), **Dr. A. Krishnamoorthy**, Associate Dean of Academics (SEEE), **Dr. K. Muthu Balan**, Associate Dean (Research), **Dr. S. Jayalalitha**, Professor (SEEE), and **Dr. M. Sridharan**, Associate Dean (Research) (SASHE), for providing me with the necessary institutional resources during the entire duration of my research.

I would like to express my heartfelt thanks and gratitude to my research supervisor, **Dr. K. Ramkumar**, Associate Dean of Academics, School of Computing (SoC). His guidance and expertise were instrumental throughout this research endeavor. I am truly grateful for the opportunity to work under his guidance, and his mentorship has significantly shaped my academic and research pursuits. His valuable insights and constructive feedback have been pivotal in shaping the direction of this thesis.

My Sincere thanks to the external and internal doctoral committee members, namely **Dr. Viswanath Kumar Ganesan**, Senior Research Scientist/TCS-Chennai, **Dr. Sivakumaran N**, National Institute of Technology, Tiruchirappalli, **Dr. Venkatesh S**, Associate Professor, School of Electrical and Electronics Engineering (SEEE), **Dr. K. Narasimhan**, Associate Professor (SEEE), and **Dr. B. Shanthi**, Professor, School of Computing (SoC).

I thank **Dr. R. Amirtharajan**, Professor (SEEE), **Dr. K. Adalarasu**, Professor (SEEE), **Dr. Subramaniyaswamy Vairavasundaram**, Professor, School of Computer Science and Engineering, VIT, Vellore, and **Dr. Ketan Kotecha**, Professor, Department of Computer Science and Engineering, Symbiosis Institute of Technology, for their valuable contributions, suggestions, and engaging discussions throughout the research process. I would like to express my sincere gratitude to my friends for their patience, understanding, and unwavering encouragement throughout the most challenging phases of this research. Their steadfast support has been instrumental in helping me persevere and navigate numerous obstacles.

I am deeply thankful to the Department of Science and Technology – Science and Engineering Research Board (DST/SERB), and to SASTRA Deemed to be University, Thanjavur, India, for their generous funding and the infrastructural support that enabled the successful execution of this work. Words fall short in expressing my profound gratitude to my parents, Mr. K. Sabapathy and Mrs. S. Vijayalakshmi. Their unconditional support, enduring sacrifices, and steadfast belief in my abilities have not only shaped my character but also instilled in me the resilience and determination necessary to pursue knowledge with purpose and optimism.

Finally, I extend my heartfelt appreciation to all those who have accompanied and

supported me on this academic journey. Your contributions, whether large or small, have been invaluable and deeply appreciated.

PRIYADHARSHINI S



**SASTRA**  
ENGINEERING · MANAGEMENT · LAW · SCIENCES · HUMANITIES · EDUCATION  
DEEMED TO BE UNIVERSITY  
(U/S 3 of the UGC Act, 1956)



THINK MERIT | THINK TRANSPARENCY | THINK SASTRA  
T H A N J A V U R | K U M B A K O N A M | C H E N N A I

## CERTIFICATE

This is to certify that the thesis titled **A Predictive Framework for Early Detection and Personalised Monitoring of Parkinson's Disease Using Artificial Intelligence and Large Language Models** is the bonafide Ph.D. work done by **Priyadharshini S** at SASTRA Deemed to be University under my guidance. This is also to certify that the thesis is the original work carried out by her, and it has not previously formed the basis for the award of any Degree, Diploma, Associateship, Fellowship, or any other similar titles of any University or Institution.

**Research Supervisor**

**Dr. K. Ramkumar,**

**Associate Dean-Academics,**

**School of Computing,**

**SASTRA Deemed to be University, Thanjavur - 613401,**

**Email: [ramkumar@eie.sastra.edu](mailto:ramkumar@eie.sastra.edu)**



**SASTRA**  
ENGINEERING · MANAGEMENT · LAW · SCIENCES · HUMANITIES · EDUCATION  
**DEEMED TO BE UNIVERSITY**  
(U/S 3 of the UGC Act, 1956)



THINK MERIT | THINK TRANSPARENCY | THINK SASTRA  
T H A N J A V U R | K U M B A K O N A M | C H E N N A I

## CANDIDATE DECLARATION

I hereby declare that the Thesis entitled, **A Predictive Framework for Early Detection and Personalised Monitoring of Parkinson's Disease Using Artificial Intelligence and Large Language Models**, is my original research work, carried out under the guidance of **Dr. K. Ramkumar**, Associate Dean-Academics, SoC. This thesis has not been submitted elsewhere for the award of any degree or diploma.

Priyadarshini S

## Table of Content

LIST OF TABLES .....	xiv
LIST OF FIGURES.....	xvi
ABSTRACT .....	xix
<b>1. INTRODUCTION</b>	
1.1 Parkinson's disease .....	1
1.2 Pathophysiology of PD .....	1
1.3 Historical Perspective and Evolution in PD .....	2
1.4 Statistical Data of PD .....	4
1.5 Clinical diagnosis of PD .....	4
1.5.1 Motor symptoms .....	6
1.5.2 Non-motor symptoms .....	6
1.5.3 Montreal Cognitive Assessment .....	7
1.5.4 Unified Parkinson's Disease Rating .....	7
1.6 Neuroimaging-based diagnosis of PD .....	7
1.6.1 Magnetic Resonance Imaging .....	8
1.6.2 Dopamine Transporter Imaging .....	9
1.7 Proteomics and Neurochemical .....	9
1.7.1 Dopamine Transporter Binding .....	10
1.7.2 CSF Proteins and Structural Changes.....	10
1.8 Progression and Diagnosis of PD .....	10
1.9 An overview of AI techniques in medical .....	11
1.9.1 ML techniques in PD diagnosis .....	12
1.9.2 DL techniques in PD diagnosis.....	12

1.9.3 XAI techniques in PD diagnosis .....	13
1.9.4 LLM in PD diagnosis .....	14
1.10 Research gaps and motivation of the thesis.....	15
1.11 Objective and specific aims .....	16
1.12 Thesis organization .....	17
<b>2. LITERATURE SURVEY</b>	
2.1 Introduction .....	18
2.2 PD diagnosis using Neuroimaging data .....	19
2.2.1 MRI Analysis.....	19
2.2.2 Iron Deposition .....	20
2.2.3 Functional Imaging with AI Models.....	20
2.2.4 Synthesis and Transition to ML.....	21
2.3 ML models for PD classify by Multimodality.....	23
2.3.1 Synthesis and Transition to DL.....	25
2.4 DL models for PD classify .....	27
2.4.1 EEG-Based DL Approaches.....	27
2.4.2 Neuroimaging and Explainable CNN .....	27
2.4.3 Clinical, CSF and Deep Learning.....	28
2.4.4 Multimodal Fusion .....	29
2.4.5 Hybrid and DL Architectures.....	29
2.4.6 Synthesis and Implications .....	30
2.5 LLMs in the medical field .....	32
2.5.1 Synthesis Linkage to Research Goals .....	34
2.6 Summary .....	34

<b>3. DIAGNOSIS OF PD BASED ON MRI SCANS</b>	
3.1 Introduction .....	37
3.2 Dataset Description .....	39
3.3 Data pre-processing .....	41
3.3.1 Brain extraction .....	42
3.3.2 Registration .....	43
3.3.3 Bias Field Correction .....	44
3.3.4 Subcortical segmentation .....	44
3.3.5 Label fusion and inverse registration .....	44
3.4 Selection of subcortical region .....	47
3.5 Radiomics-based Classification of PD .....	49
3.6 ML classifier .....	51
3.7 XAI Techniques .....	51
3.8 Result and Analysis of ML classifiers .....	53
3.9 Performance Evaluation and Analysis .....	55
3.10 Analysis and Interpretation using XAI .....	58
3.11 Statistical Analysis .....	61
3.12 Discussion on Radiomic and XAI Insights .....	63
3.13 Summary .....	64
<b>4. HYBRID DL FRAMEWORK FOR PD DETECTION</b>	
4.1 Introduction .....	66
4.2 Dataset Collection and Pre-Processing .....	68
4.3 Deep learning architecture .....	69
4.3.1 Proposed 3D-CNN .....	70
4.3.2 Improved 3D-ResNet .....	70

4.4 Feature Extraction .....	72
4.5 Feature Fusion Techniques .....	74
4.5.1 Feature Concatenation .....	74
4.5.2 Canonical Correlation Analysis .....	75
4.6 Feature optimization .....	77
4.6.1 Mathematical model of WOA .....	77
4.6.2 WOA for feature selection .....	79
4.7 Results and Analysis .....	80
4.7.1 Proposed 3D-CNN architecture .....	81
4.7.2 Model-2 (Improved 3D-ResNet) .....	85
4.7.3 Fusion without Optimization .....	86
4.7.4 Fusion with WOA optimization.....	87
4.8 Discussion .....	88
4.9 Summary .....	91
<b>5. ARTIFICIAL FRAMEWORK FOR PD ANALYSIS</b>	
5.1 Introduction .....	94
5.2 Data Collection .....	96
5.3 Delineation of multimodal data features .....	97
5.3.1 Neuroimaging data .....	97
5.3.2 SBR values from SPECT .....	97
5.3.3 CSF protein biomarkers .....	98
5.3.4 Clinical data .....	98
5.4 Proposed Methodology .....	100
5.4.1 Proposed 1D-CNN model .....	101
5.4.2 Fine-tuned LLM for PD analysis.....	103

5.4.3 Cloud-based interactive system .....	104
5.5 Results and Analysis .....	105
5.5.1 Statistical Analysis of Multimodal .....	106
5.5.2 Performance evaluation of 1D-CNN.....	106
5.5.3 LLM-based clinical diagnosis .....	108
5.5.4 LLM-Driven medical report .....	110
5.5.5 Integrated Cloud-Based Management.....	113
5.6 Discussion .....	116
5.6.1 Datatypes for finetuning and Testing.....	116
5.6.2 Parameter Fine-Tuning LLM .....	117
5.6.3 Evaluation for finetune model.....	117
5.6.3.1 Structured Prompt-Pairs.....	118
5.6.3.2 Evaluating using GPT-4.0.....	118
5.6.3.3 Three-step Automated.....	119
5.7 Summary .....	119
<b>6. CONCLUSION AND FUTURE WORK</b>	
6.1 Conclusion .....	121
6.2 Scope of Future work .....	123
<b>REFERENCES .....</b>	125
<b>APPENDIX A: LIST OF ABBREVIATIONS .....</b>	139
<b>APPENDIX B: LIST OF PUBLICATIONS .....</b>	143

## List of Tables

Table 2.1	Summarises the diagnosis of PD based on neuroimaging techniques using AI approaches .....	22
Table 2.2	Summary of ML Methods for PD Detection .....	25
Table 2.3	Summary of DL Methods for PD Detection .....	30
Table 3.1	Pseudocode of the MRI pre-processing pipeline .....	45
Table 3.2	Labels, Structures, and Uses of each subcortical structure in the Atlas .....	48
Table 3.3	Explanation of feature type and number of features extracted from ROI .....	50
Table 3.4	Specifications used for different ML models .....	54
Table 3.5	Performance of the model before and after applying the up-sampling technique .....	54
Table 3.6	Comparison of different machine learning classifiers .....	56
Table 3.7	Statistical analysis of the selected features .....	62
Table 4.1	A Comparative Overview of Performance and Image Quality for Different MRI Machines .....	69
Table 4.2	Algorithm for Feature Concatenation .....	76
Table 4.3	Ablation Study of the Constructed 3D-CNN Architecture.....	83
Table 4.4	Comparative analysis of PD prediction performance using different architectural models and Feature Extraction Methods .....	87
Table 4.5	Performance analysis of models 3 and 4 .....	87
Table 4.6	Parameter values for all algorithms .....	89
Table 5.1	Details of the multimodal data with p-values of the biomarkers used in the current research .....	99
Table 5.2	Comparison of features for multiclass classification in PD.....	109
Table 5.3	PD-relevant data domains Q/A pairs for fine-tuned and testing the model .....	116

Table 5.4	Sample JSON format .....	118
Table 5.5	Evaluation criteria for the generated response .....	119

## List of Figures

<p>Figure 1.1 Overview of PD pathogenesis involving iron overload, oxidative stress, alpha-synuclein accumulation, and mitochondrial dysfunction leading to ferroptosis .....</p> <p>Figure 1.2 Projected global distribution of PD cases, emphasising the significant growth in Asia by 2030 .....</p> <p>Figure 1.3 Symptoms of PD .....</p> <p>Figure 1.4 Stages of PD highlighting progression of motor and non-motor symptoms over time .....</p> <p>Figure 1.5 Multimodal input features for Parkinson's Disease detection categorized by AI pipelines. ....</p> <p>Figure 1.6 Taxonomy of XAI techniques based on stage, scope, and model dependency .....</p> <p>Figure 1.7 XAI workflow showing data flow from various input types to explainable outputs via a black-box model .....</p> <p>Figure 2.1 Evolution of AI methods in PD diagnosis, highlighting data modalities, techniques, clinical applications, key challenges, and the proposed thesis framework. ....</p> <p>Figure 3.1 Workflow of the designed system integrating XAI and ML approaches for early diagnosis and progression tracking of PD.....</p> <p>Figure 3.2 NIfTI image from PPMI database .....</p> <p>Figure 3.3 Detailed Analysis of the pre-processing of MRI scans ....</p> <p>Figure 3.4 Label Fusion and Extracted Voxels from an MRI after Segmentation .....</p> <p>Figure 3.5 Overview of the complete workflow designed for PD grading.....</p> <p>Figure 3.6 Scatter plot visualising feature distribution before and after applying SMOTE, highlighting class separation improvement.....</p> <p>Figure 3.7 (a) Accuracy of GB classifier across 5-Fold Cross-Validation splits, (b) Confusion matrix for GB classifier, (c) ROC curve three different classes for GB Classifiers model.....</p>	<p style="text-align: right;">3</p> <p style="text-align: right;">5</p> <p style="text-align: right;">5</p> <p style="text-align: right;">11</p> <p style="text-align: right;">13</p> <p style="text-align: right;">15</p> <p style="text-align: right;">15</p> <p style="text-align: right;">35</p> <p style="text-align: right;">39</p> <p style="text-align: right;">41</p> <p style="text-align: right;">42</p> <p style="text-align: right;">46</p> <p style="text-align: right;">52</p> <p style="text-align: right;">55</p> <p style="text-align: right;">58</p>
--	---

Figure 3.8	Interpretation of the GB ML classifier for classification of PD disease (a) SHAPASH represent the local feature and its feature importance to individual instances (b) LIME shows the local importance of the feature by single instances (c) SHAP shows global and local feature importance and produce the values for every individual feature.....	61
Figure 4.1	Overall architecture of the proposed method .....	71
Figure 4.2	Proposed 3D-CNN .....	72
Figure 4.3	Improved 3D-ResNet .....	73
Figure 4.4	Feature Selection Process Using WOA .....	82
Figure 4.5	The overall performance comparison of algorithms across different metrics (a) Accuracy, (b) Average Fitness, (c) Best Fitness, (d) Selection size, (e) Time .....	90
Figure 4.6	Convergence curves generated using various algorithms for the PPMI dataset .....	91
Figure 4.7	Comparison of performance metrics for the proposed method.....	92
Figure 5.1	Detailed architecture of the classification (1D-CNN) layer .....	103
Figure 5.2	Finetuned LLM model for diagnosis of PD .....	104
Figure 5.3	The overview workflow of the 1D-CNN classifier model	108
Figure 5.4	Image-based example cases and corresponding Fine-tuned LLM responses.....	111
Figure 5.5	Text-based examples cases and their corresponding finetuned model responses .....	111
Figure 5.6	Report generated by the fine-tuned LLM using a standardized diagnostic template .....	113
Figure 5.7	User interface of the proposed cloud-based platform (a) Proteins prediction module interface, allowing users to input clinical data for predicting protein levels related to PD. (b) Brain imaging data analysis interface, supporting the upload and processing of NiFTI files for advanced analysis. (c) Parkinson Chat Bot interface, acting as a digital assistant specialized in PD. (d) User navigation	

dropdown showcasing available tools within the platform to access various functionalities.....	115
Figure 5.8 Brain imaging data analysis interface, supporting the upload and processing of NiFTI files for advanced analysis .....	115

## Abstract

A Predictive Framework for Early Detection and Personalised Monitoring of Parkinson's Disease Using Artificial Intelligence and Large Language Models

Priyadharshini S

Dr K Ramkumar, Ph.D

Parkinson's Disease (PD) is a multifaceted and progressive neurodegenerative disorder that presents a spectrum of motor and non-motor symptoms. Early and accurate diagnosis is essential for effective disease management and improved patient outcomes, yet remains clinically challenging due to symptom overlap and diagnostic limitations. This thesis proposes a comprehensive and interpretable artificial intelligence (AI)-driven diagnostic framework that aims to transform the early detection, personalised monitoring, and treatment recommendation process for PD. The proposed solution integrates deep learning, radiomics, evolutionary optimisation, and large language models (LLMs), ensuring a highly accurate and clinically adaptable system.

The research begins by analysing T2-weighted 3D Magnetic Resonance Imaging (MRI) scans sourced from the Parkinson's Progression Marker Initiative (PPMI) database. A robust preprocessing pipeline comprising brain extraction, registration, bias correction, normalization, and segmentation is applied. From the segmented subcortical brain regions, 107 radiomics features are extracted, of which the top 20 most predictive are selected using Pearson correlation, recursive feature elimination, and ranking techniques. Statistical validation is conducted using ANOVA, pairwise t-tests, and Kruskal-Wallis H-tests. Multiple machine learning algorithms are evaluated, and the Gradient Boosting (GB) model, enhanced by the Synthetic Minority Oversampling Technique (SMOTE), attains an improved diagnostic accuracy of 96.8 % up from 86 %. To enhance transparency and clinical trust, Explainable AI (XAI) methods such as,

SHAP and LIME are implemented, offering interpretable visual insights into model predictions.

For advanced volumetric analysis, a custom 3D Convolutional Neural Network (3D-CNN) is designed and optimised through architectural refinement and hyperparameter tuning, achieving an accuracy of 93.4%. This model outperforms the baseline and complements an existing 3D-ResNet, which independently achieves 90% accuracy. Canonical Correlation Analysis (CCA) is then employed to fuse high-level features from both networks, yielding a combined accuracy of 95%. Further enhancement is achieved through the application of the Whale Optimisation Algorithm (WOA), a biologically inspired evolutionary technique, which boosts the final classification accuracy to 97%.

Recognising the multi-dimensional nature of PD, the thesis expands into multimodal data integration, encompassing MRI, SPECT scans, cerebrospinal fluid (CSF) protein biomarkers, and clinical scores. A 1D-CNN model is developed using 121 multimodal features and initially achieves an accuracy of 94.9%. With the inclusion of biologically derived ratio-based biomarkers, this accuracy increases to 96.9%. The integration of a fine-tuned ChatGPT-4.0 Mini model bridges AI-driven insights with clinical narratives, enabling personalised report generation, improved patient engagement, and real-time clinical decision support. A cloud-based platform is developed to enable scalable deployment with features like real-time inference, chatbot-assisted communication, and automated medical summaries.

Overall, this thesis presents a unified, explainable, and clinically deployable AI framework that significantly enhances the capabilities of PD diagnosis and personalized care. By integrating deep learning, radiomics, evolutionary optimization, and Large

Language Models within a cloud-enabled platform, the proposed system establishes a novel benchmark for future clinical AI applications in the management of neurodegenerative disease.



## CHAPTER 1: INTRODUCTION

### 1.1 Parkinson's disease

Parkinson's Disease (PD) is the second most prevalent neurodegenerative disorder globally, marked by the gradual degeneration of dopaminergic neurons within the substantia nigra (SN) pars compacta, accompanied by pathological accumulation of alpha-synuclein protein aggregates known as Lewy bodies [1], [2]. Clinically, Parkinson's is identified through prominent motor symptoms including bradykinesia, resting tremor, muscular rigidity, and impaired postural reflexes. However, the complexity of Parkinson's extends beyond these motor signs, encompassing numerous non-motor symptoms such as cognitive impairment, mood disturbances, sleep disorders, and autonomic dysfunctions [3]–[5]. Importantly, these non-motor symptoms often precede motor manifestations by several years, highlighting the need for earlier and more accurate diagnostic tools.

### 1.2 Pathophysiology of PD

At a cellular and molecular level, the progression of Parkinson's involves a complex interplay of biological disruptions, notably oxidative stress, mitochondrial dysfunction, dysregulation of iron metabolism, and persistent neuroinflammation. Elevated iron levels participate in the Fenton reaction, increasing the production of reactive oxygen species (ROS), thereby exacerbating oxidative stress and cellular injury. Simultaneously, dysfunctional mitochondrial dynamics marked by impaired fission and fusion processes involving proteins like DRP1 contribute to bioenergetic failure in neurons. These disruptions are closely linked with alpha-synuclein aggregation, leading to the formation of Lewy bodies, which further compromise neuronal health by propagating inflammatory responses. Collectively, these pathological mechanisms lead to ferroptosis, an iron-dependent form of programmed cell death, perpetuating neuronal degeneration [6]–[8]. As shown in Figure 1.1, the interplay between iron overload, ROS generation, mitochondrial dysfunction, and alpha-synuclein accumulation establishes a

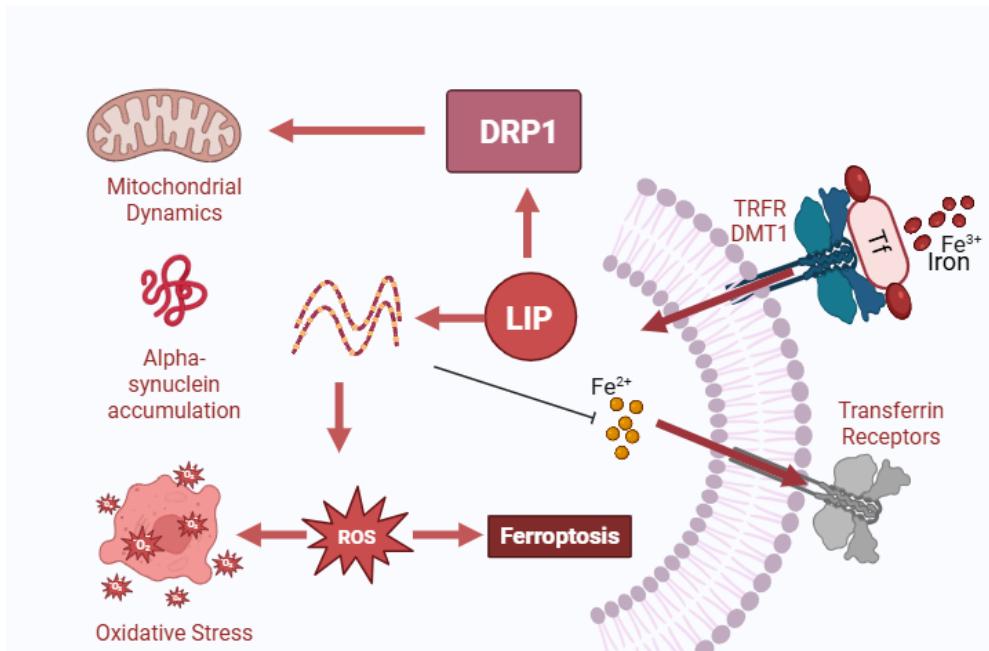
self-sustaining pathological cycle, highlighting critical targets for therapeutic intervention aimed at slowing or halting disease progression. Despite considerable advances in understanding these mechanisms, current diagnostic practices predominantly rely on motor symptom manifestation, typically indicating substantial neuronal loss has already occurred. This limitation highlights the urgent need for integrative, data-driven diagnostic approaches, especially leveraging artificial intelligence (AI), to facilitate earlier disease detection and intervention.

### 1.3 Historical Perspective and Evolution in PD Research

The understanding of PD has evolved significantly from its initial clinical descriptions to sophisticated insights into its molecular pathology and advanced therapeutic strategies. The earliest clinical characterisation was presented by James Parkinson in 1817, describing the condition as "shaking palsy" [9]. This foundational observation paved the way for future explorations, leading to notable scientific breakthroughs. By the early 20th century, the identification of neuropathological changes within the SN offered the first concrete evidence linking brain structure alterations to Parkinson's symptoms. A critical milestone occurred in the mid-20th century with the discovery of dopamine's crucial role in motor function by Arvid Carlsson, catalysing the development of Levodopa therapy [10]. Although Levodopa does not cure PD, it significantly improves symptom management and patient quality of life.

Subsequently, the 1990s witnessed another major advancement with the introduction of Deep Brain Stimulation (DBS), particularly targeting the subthalamic nucleus, offering substantial symptom relief in advanced stages where medication alone was insufficient [11]. Alongside these therapeutic developments, genetic research unveiled mutations in alpha-synuclein linked to familial Parkinson's, shifting diagnostic and therapeutic approaches from purely symptomatic management towards biomarker-driven strategies. Advancements in neuroimaging further enhanced diagnostic capabilities, enabling earlier and more precise visualisation of structural and functional brain changes through techniques like Magnetic

Resonance Imaging (MRI) and Dopamine Transporter Imaging (DaTscan) [12]. In recent years, AI and machine learning (ML) technologies have begun to transform diagnostic practices.



**Figure 1.1** Overview of PD pathogenesis involving iron overload, oxidative stress, alpha-synuclein accumulation, and mitochondrial dysfunction leading to ferroptosis.

#### 1.4 Statistical Data on PD

The global burden of PD has escalated at an alarming pace over the past few decades, making it not only a growing neurological challenge but also a critical public health concern. As life expectancy increases and aging populations expand, particularly in developing regions, PD cases are projected to rise significantly in both absolute and proportional terms. In 2005, approximately 4.1 million individuals were living with PD worldwide. This number is expected to more than double, reaching an estimated 8.7 million by 2030. The demographic shift is especially pronounced in Asia, where countries like China and India are poised to experience the steepest increase in PD prevalence. China alone is projected to account for nearly 57% of all global cases by 2030, up from 48% in 2005. India's contribution, while holding steady at

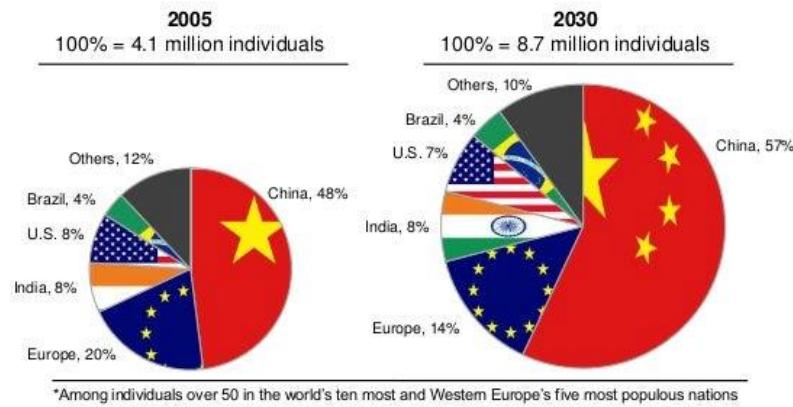
8%, represents a substantial absolute increase, given its population size and increasing longevity [13], [14]. In contrast, regions such as Europe and North America are expected to witness a relative decline in their share of global PD cases from 20% to 14% in Europe, and from 8% to 7% in the United States despite stable or modest increases in actual case numbers. These shifts reflect not just demographic transitions, but also disparities in healthcare access, early diagnostic infrastructure, and public awareness campaigns. As illustrated in Figure 1.2, this geographic redistribution highlights the urgent need for tailored diagnostic tools and healthcare planning. Countries with rapidly ageing populations, but limited specialised neurological services, may face a disproportionate burden. This underscores the importance of scalable, cost-effective, and AI-assisted diagnostic frameworks that can bridge the gap between resource availability and growing clinical demand. Ultimately, understanding the evolving epidemiology of PD is critical not only to anticipate healthcare needs but also to inform the development of predictive and preventative strategies that can be adapted to regional contexts and demographic realities.

## 1.5 Clinical diagnosis of PD

PD presents with a diverse and progressive range of symptoms that affect both motor and non-motor systems. While clinical diagnosis has traditionally centred on motor abnormalities, a broader understanding now recognises that non-motor symptoms may precede visible motor signs by several years. This evolution in clinical insight has redefined how PD is detected and managed across its continuum. Figure 1.3 visually illustrates the multi-systemic nature of PD, highlighting how symptoms span the nervous system, gastrointestinal tract, musculoskeletal system, and cognitive-emotional domains. These wide-ranging manifestations require a comprehensive clinical evaluation grounded in structured neurological examination and validated rating scales to achieve a timely and accurate diagnosis.

## The burden of Parkinson disease and other neurodegenerative conditions is growing

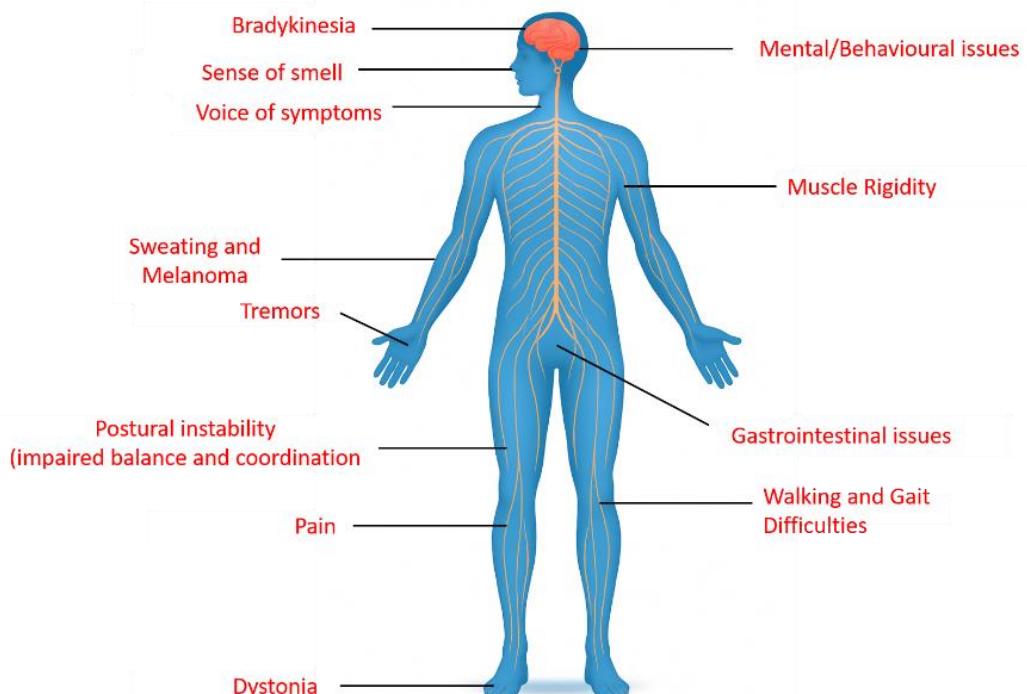
Distribution of individuals with Parkinson disease by country from 2005 to 2030\*



Source: *Neurology* 2007;68:384-6

5

**Figure 1.2:** Projected global distribution of PD cases, emphasising the significant growth in Asia by 2030.



**Figure 1.3:** Symptoms of PD

### 1.5.1 Motor symptoms

Motor symptoms remain the diagnostic cornerstone of PD and are often the most recognizable clinical signs. As illustrated in Figure 1.3, these include:

- **Bradykinesia:** Slowness of movement, most evident during initiation of voluntary actions such as walking or writing.
- **Tremors:** Typically manifesting at rest, often beginning unilaterally in the hand and progressing to other limbs.
- **Rigidity:** Increased muscle tone leading to stiffness and reduced range of motion.
- **Postural Instability:** Impaired balance and coordination, contributing to falls.
- **Gait Abnormalities and Dystonia:** Shuffling steps, stooped posture, and involuntary muscle contractions affecting limb positioning. These symptoms often present asymmetrically and evolve with disease progression. Their pattern, onset, and response to dopaminergic treatment aid in differentiating idiopathic PD from other parkinsonian syndromes.

### 1.5.2 Non-motor symptoms

Non-motor symptoms, although less conspicuous, significantly impact quality of life and often precede motor onset. Figure 1.3 outlines several key non-motor features:

- **Cognitive Dysfunction and Mental/Behavioural Changes:** Including memory deficits, depression, anxiety, and apathy.
- **Loss of Smell (Hyposmia):** A common prodromal symptom, often unnoticed or misattributed.
- **Sleep Disturbances:** Such as REM sleep behavior disorder, which may signal early neurodegeneration.

- **Autonomic Dysfunction:** Including gastrointestinal issues (e.g., constipation), urinary urgency, and excessive sweating.
- **Vocal Changes and Fatigue:** Such as reduced voice volume and energy levels.
- **Pain and Sensory Disturbances:** Unexplained musculoskeletal or neuropathic pain may accompany PD even before motor signs emerge. These symptoms highlight the importance of a holistic diagnostic approach that goes beyond visible impairments to recognize underlying physiological dysfunctions.

### **1.5.3 Montreal Cognitive Assessment (MoCA)**

Given the cognitive impairments associated with Parkinson's, the Montreal Cognitive Assessment (MoCA) has become an essential screening tool. MoCA evaluates multiple cognitive domains, including executive function, memory, language, and visuospatial skills. Due to its high sensitivity for detecting mild cognitive impairment and its brevity, MoCA is widely adopted in both clinical and research settings. Typically, scores below 26 out of 30 suggest cognitive impairment [5].

### **1.5.4 Unified Parkinson's Disease Rating Scale (UPDRS)**

The Unified Parkinson's Disease Rating Scale (UPDRS), particularly its revised MDS-UPDRS version, systematically assesses disease severity and progression. It comprises four parts: non-motor experiences of daily living, motor experiences of daily living, clinician-rated motor examination, and motor complications. UPDRS facilitates accurate symptom quantification, monitors therapeutic responses, and guides clinical management decisions, making it invaluable for clinical and research applications [10].

## **1.6 Neuroimaging-based diagnosis of PD**

While clinical symptoms remain central to PD diagnosis, advances in neuroimaging have introduced objective, non-invasive tools that provide deeper insights into the brain's structural

and functional changes. Imaging techniques have become increasingly valuable for differentiating PD from other Parkinsonian syndromes, identifying early-stage pathology, and supporting clinical decision-making. Modern neuroimaging complements traditional assessments by visualising dopaminergic neuron integrity and monitoring disease progression. Among the most widely used modalities are MRI and DaTscan. These approaches allow clinicians to evaluate structural degeneration and functional dopaminergic deficits, often before overt motor symptoms appear.

### **1.6.1 Magnetic Resonance Imaging (MRI)**

MRI is a versatile and widely available imaging technique that plays a dual role in PD assessment: excluding other structural causes of parkinsonism (e.g., vascular lesions or tumors) and detecting subtle neuroanatomical changes associated with PD. Although conventional MRI is limited in visualising early dopaminergic degeneration, advances in high-field imaging and specialised sequences have improved its diagnostic relevance. In PD, particular focus is placed on deep brain structures such as the SN, globus pallidus, subthalamic nucleus, and red nucleus. Techniques like T2-weighted and T2-weighted\* imaging help visualise iron deposition, a pathological hallmark of PD in the SN [15], [16]. Increased susceptibility due to iron accumulation results in signal loss in these regions, providing indirect evidence of dopaminergic neuronal loss. Additionally, diffusion tensor imaging (DTI) and magnetisation transfer imaging have been used to evaluate microstructural integrity and tissue contrast in affected regions. These modalities have shown potential in detecting early disease changes, although they are not yet standard in routine clinical practice. While MRI cannot definitively diagnose PD, its ability to detect characteristic atrophy patterns and rule out PD mimics enhances diagnostic confidence, especially when combined with clinical and biochemical findings.

### **1.6.2 DaTscan (Dopamine Transporter Imaging)**

DaTscan, a form of single-photon emission computed tomography (SPECT), allows for the visualisation of dopamine transporter availability in the striatum, a critical region affected in PD. It involves the intravenous injection of a radiotracer, typically I-123 ioflupane, which binds selectively to presynaptic DAT. In PD, degeneration of dopaminergic neurons in the SN leads to a reduction in DAT binding, especially in the posterior putamen. This reduction manifests as asymmetric tracer uptake in DaTscan images, serving as a reliable biomarker of presynaptic dopaminergic dysfunction [17]. The degree of reduction often correlates with disease severity and symptom lateralization. DaTscan is particularly useful for distinguishing PD from non-degenerative conditions such as essential tremor, drug-induced parkinsonism, and vascular parkinsonism, where DAT binding remains largely intact. However, its diagnostic specificity is limited when differentiating PD from atypical parkinsonian syndromes like multiple system atrophy (MSA) or progressive supranuclear palsy (PSP), as all exhibit similar reductions in DAT density. Despite these limitations, DaTscan remains one of the most clinically accepted tools for functional imaging in PD and is often used as a supportive diagnostic measure in complex or uncertain cases.

### **1.7 Proteomics and Neurochemical-Based Diagnosis of PD**

Recent advancements in molecular and neurochemical biomarkers have considerably enriched the diagnostic landscape of PD. Unlike traditional clinical assessments reliant on visible symptoms, these biomarkers offer objective, quantifiable insights into the disease's underlying pathological mechanisms, often manifesting before motor symptoms become evident. Two primary biomarker categories have gained prominence: Dopamine Transporter Binding Ratios (SBR values) measured through imaging techniques, and protein biomarkers present in cerebrospinal fluid (CSF), complemented by structural brain assessments through imaging.

### **1.7.1 Dopamine Transporter Binding Ratios (SBR Values)**

In PD, these SBR ratios exhibit significant reductions, notably within the posterior putamen, reflecting the loss of presynaptic dopaminergic neurons. Such decreases often present asymmetrically and correlate with the severity and lateralization of clinical symptoms. Hence, SBR values serve as critical, non-invasive biomarkers, providing real-time insights into dopaminergic system integrity, essential for early diagnosis and progression monitoring.

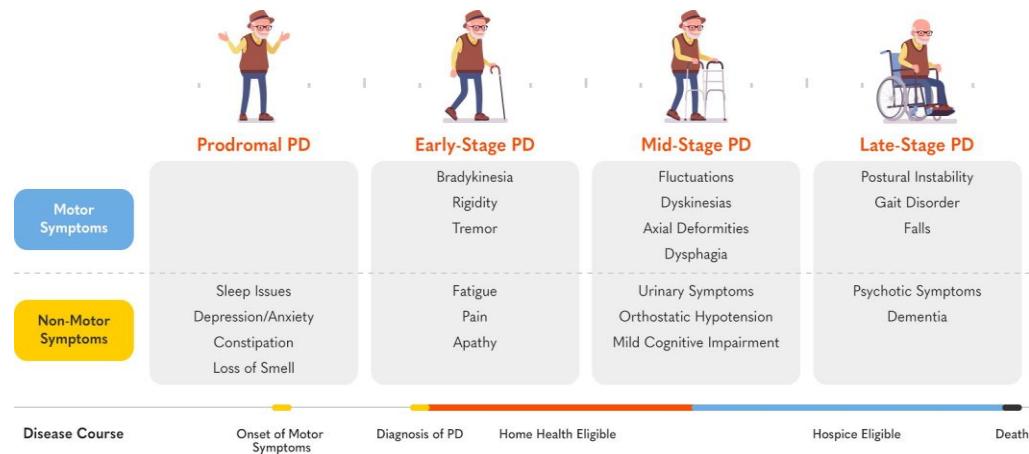
### **1.7.2 Cerebrospinal Fluid Proteins and Structural Brain Changes**

Biochemical analysis of CSF has revealed distinctive profiles of key proteins such as alpha-synuclein, total tau (tTau), phosphorylated tau (pTau), and amyloid-beta 1–42 (A $\beta$ 1–42). In patients with PD, decreased levels of alpha-synuclein and A $\beta$ 1–42, alongside altered tau protein levels, signal ongoing synaptic dysfunction and neuronal degeneration even in early disease stages. Concurrently, structural imaging via MRI identifies critical morphological changes, including white matter connectivity disruptions and gray matter atrophy, particularly affecting regions governing motor and cognitive functions. Integrating CSF proteomics with structural imaging enhances diagnostic precision and helps differentiate PD from other neurodegenerative disorders exhibiting overlapping symptom profiles.

## **1.8 Progression and Diagnosis of PD**

Despite advancements in Parkinson's disease diagnosis, around 5% of patients in the U.S. are still misdiagnosed each year. At the primary care level, the situation is even more concerning—nearly 1 in 5 patients with serious conditions may receive an incorrect diagnosis. [18]. Clinically, PD progression typically encompasses four stages: prodromal, early, mid, and late-stage, each presenting unique diagnostic challenges as shown in Figure 1.4. The prodromal phase is especially significant, characterized by subtle non-motor symptoms such as sleep disturbances, constipation, mood changes, anxiety, and olfactory impairment, often manifesting years prior to classical motor signs. PD progresses to the early and mid-stages, with motor

symptoms like bradykinesia, rigidity, tremors, motor fluctuations, and dyskinesias. Cognitive impairments and autonomic dysfunctions, such as orthostatic hypotension and mild cognitive impairment, further complicate management strategies. In advanced stages, patients frequently experience severe postural instability, gait disorders, and prominent neuropsychiatric manifestations, including dementia and psychosis. Effective diagnosis across these stages requires a comprehensive approach, integrating clinical evaluations, structured assessment tools, and objective biomarkers such as neuroimaging and CSF analysis.



**Figure 1.4:** Stages of PD highlighting progression of motor and non-motor symptoms over time

## 1.9 An overview of AI techniques in medical field

AI has significantly transformed healthcare, enabling precise, scalable, and data-driven solutions for diagnosis, prognosis, and therapeutic planning. Within complex neurological disorders like PD, AI methodologies facilitate the extraction of intricate patterns from diverse clinical, imaging, and molecular datasets, surpassing human perceptual capabilities.

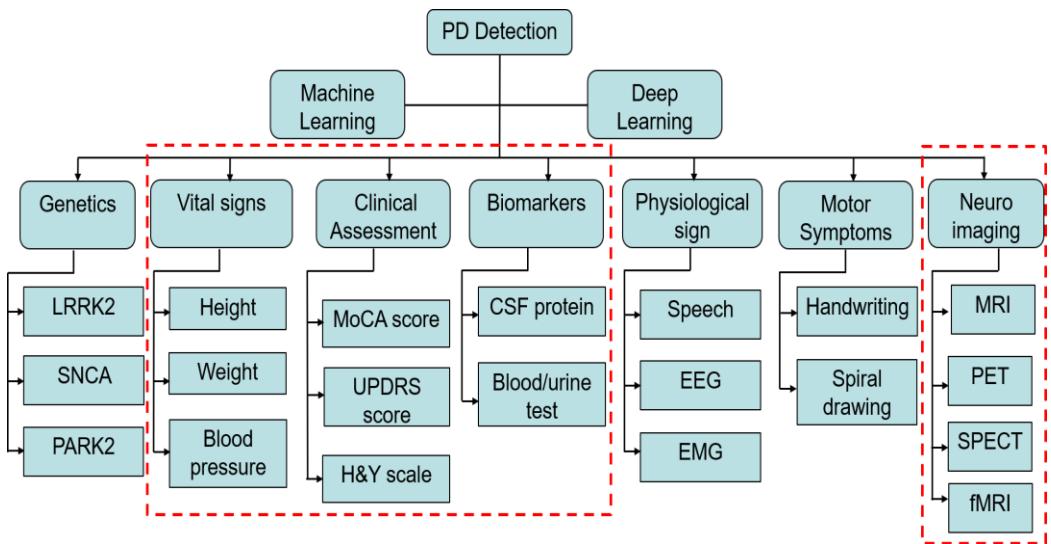
### **1.9.1 ML techniques in PD diagnosis**

Machine Learning (ML) is a subset of AI that utilises structured datasets to make predictive or diagnostic assessments by identifying patterns inherent within data. ML has proven valuable in diagnosing PD by distinguishing subtle patterns in clinical symptoms, neuroimaging, biochemical markers, and biosignals, which traditional clinical methods might overlook. ML algorithms such as supervised, unsupervised, semi-supervised, and reinforcement learning have enhanced diagnostic accuracy, improved disease staging, and facilitated personalised treatment strategies [19], [20]. The reliability and efficiency of these models heavily depend on dataset quality and volume, making preprocessing and feature selection pivotal for optimal performance. Previous research [21] extensively reviewed ML applications in PD, highlighting significant improvements in early detection, accurate classification, and clinical decision support.

### **1.9.2 DL techniques in PD diagnosis**

Deep Learning (DL) is a subset of ML, and is especially adept at processing high-dimensional, unstructured data by learning layered representations. DL techniques like CNNs, Recurrent Neural Networks (RNNs), Deep Belief Networks (DBNs), and Deep Boltzmann Machines have shown marked improvements in analysing medical imagery and time-series biosignals for PD diagnosis [22], [23]. Unlike traditional ML methods, DL architectures eliminate the need for manual feature extraction by learning directly from raw data. Research has highlighted the enhanced capability of DL models in detecting structural and functional changes from neuroimaging and physiological datasets [24]. These models support early and accurate diagnosis, even in cases with minimal clinical manifestations. Figure 1.5 illustrates the diverse data modalities involved in PD detection, spanning genetics (e.g., LRRK2, SNCA, PARK2), vital signs (e.g., height, weight, blood pressure), clinical assessments (e.g., MoCA, UPDRS, H&Y scale), biomarkers (e.g., CSF proteins), physiological signals (e.g., EEG, EMG), motor symptoms, and neuroimaging data (e.g., MRI, PET, SPECT, fMRI). The red-dotted boxes

highlight the input data streams most frequently explored in recent studies for early PD diagnosis using ML and DL techniques.



**Figure 1.5** Multimodal input features for Parkinson's Disease detection categorized by AI pipelines.

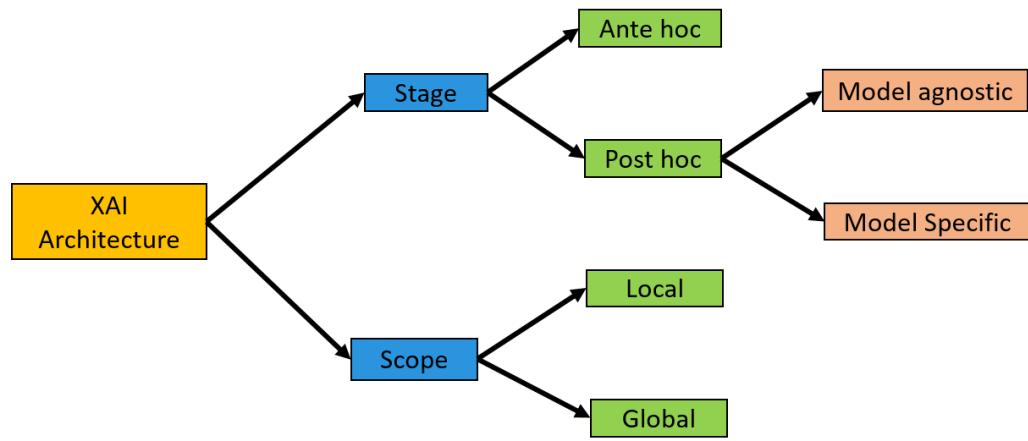
### 1.9.3 XAI techniques in PD diagnosis

Explainable Artificial Intelligence (XAI) has become a critical field aimed at enhancing the transparency, interpretability, and clinical integration of AI models, especially within healthcare contexts such as PD diagnosis. Deep learning models, while powerful, often operate as "black boxes," making it difficult for clinicians to interpret the reasoning behind their predictions. This issue underscores the necessity for robust XAI methods. Post-hoc interpretability techniques such as Local Interpretable Model-Agnostic Explanations (LIME), Shapley Additive exPlanations (SHAP) and Gradient-weighted Class Activation Maps (Grad-CAM) have been widely applied to clarify predictions from complex models. These techniques provide insights into model behaviours and prediction rationales, promoting clinical accountability and enhancing trust among healthcare professionals [25]–[27].

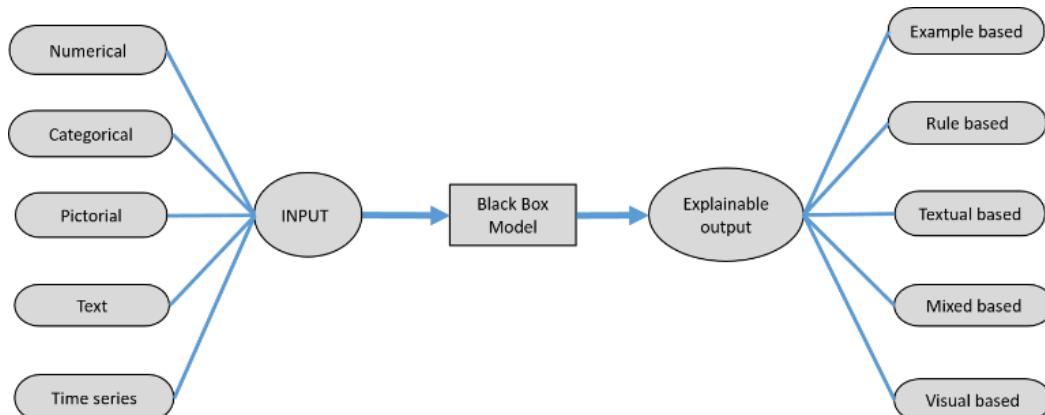
Figure 1.6 categorizes the XAI techniques that can be systematically categorized based on three dimensions: the stage at which interpretation occurs (ante hoc or post hoc), the scope of explanation (local or global), and their model dependency (model-agnostic or model-specific). Ante hoc approaches embed interpretability into the model during development, whereas post hoc techniques derive explanations after model training, making them especially valuable for black-box models commonly used in clinical AI. [28]. Additionally, model-agnostic techniques provide flexibility by being applicable across diverse model types, while model-specific methods are tailored for particular architectures or learning paradigms [29]. Figure 1.7 outlines a typical XAI workflow, demonstrating how diverse data types including numerical, categorical, pictorial, textual, and time series inputs are processed through a black-box AI model to generate explainable outputs. These outputs may take various forms such as example-based, rule-based, textual, visual, or mixed representations, thereby enabling clinicians to better interpret AI-driven decisions. In PD diagnosis, visual explanations (e.g., Grad-CAM heatmaps) highlight brain regions significant in determining diagnosis, numerical methods like SHAP quantify feature contributions, and rule-based methods elucidate the decision-making logic of diagnostic predictions. Collectively, these approaches not only facilitate informed clinical decisions but also bolster clinician confidence in AI-driven diagnostic systems.

#### **1.9.4 LLM in PD diagnosis**

Large Language Models (LLMs) like GPT-4 and BERT have redefined medical AI by effectively processing and interpreting clinical texts and patient-reported data. Their applications extend to medical documentation, clinical note summarisation, and patient interaction, particularly beneficial for remote diagnostics and healthcare delivery in resource-constrained environments [30]–[32]. Despite concerns over factual inaccuracies and model consistency, domain-specific adaptations (e.g., BioBERT, PubMedGPT) have significantly improved the clinical applicability of LLM.



**Figure 1.6** Taxonomy of XAI techniques based on stage, scope, and model dependency



**Figure 1.7** XAI workflow showing data flow from various input types to explainable outputs via a black-box model.

## 1.5 Research gaps and motivation of the thesis

Despite advancements in AI-driven diagnostics, critical research gaps persist, particularly in PD's early-stage diagnosis accuracy, multimodal data integration, and clinical interpretability. Current diagnostic frameworks predominantly rely on evident motor symptoms, often resulting

in late diagnosis after substantial neuronal damage has occurred. The integration of multimodal datasets, encompassing clinical, biochemical, and imaging data, remains insufficiently explored, limiting comprehensive diagnostic capabilities. Furthermore, existing AI models frequently lack interpretability, creating barriers to clinical adoption due to the opacity of decision-making processes. Addressing these gaps provides robust motivation for the current research. This thesis aims to develop an integrative, interpretable AI framework for early PD detection and personalized patient monitoring, leveraging multimodal data fusion and advanced XAI methodologies. Enhancing early diagnostic accuracy and model transparency not only improves clinical outcomes but also fosters greater clinician trust and patient acceptance, ultimately advancing precision medicine for PD.

## **1.6 Objective and specific aims**

The primary goal of this research is to design and develop an AI-powered Clinical Decision Support System (CDSS) that facilitates the early detection and personalised monitoring of PD. By integrating diverse modalities, including neuroimaging, clinical assessments, and biochemical markers with AI techniques, the proposed system aims to enhance diagnostic accuracy, improve interpretability, and support individualised care strategies. The specific aims of this research are:

**Specific Aim # 1:** The goal is to build a predictive system that can support early detection and track the progression of Parkinson's Disease, leveraging machine learning and explainable AI to ensure transparency and trust in the outcomes.

**Specific Aim # 2:** To design a comprehensive pipeline for MRI-based feature extraction, selection, and optimization to enhance diagnostic performance.

**Specific Aim # 3:** To implement a cloud-based clinical decision support platform that integrates Large Language Models to provide personalised diagnostic insights and treatment planning.

## 1.7 Thesis organisation

The thesis organisation is given as follows:

**Chapter 1** provides the background and motivation for PD diagnosis, outlines the key research gaps, and emphasizes the need for intelligent, computer-aided diagnostic solutions.

**Chapter 2** delivers a comprehensive literature review on PD prediction using various machine learning, deep learning, and explainable AI (XAI) techniques.

**Chapter 3** introduces an XAI-driven framework for the early detection and classification of PD using MRI-based radiomic features extracted from subcortical brain regions. This chapter proposes an integrated framework combining radiomics, machine learning, and XAI to develop interpretable and personalised diagnostic tools for neuroimaging-based PD detection.

**Chapter 4** focuses on the development and validation of custom 3d-CNN and enhanced 3d-ResNet architectures. It evaluates the performance of different models and presents a feature fusion strategy combined with feature optimisation for improved PD prediction.

**Chapter 5** presents a comprehensive AI-powered diagnostic framework that integrates multimodal data sources into a custom 1d-CNN for accurate multiclass classification. This chapter also introduces a fine-tuned large language model and a cloud-based platform to enhance model interpretability, support real-time query resolution, and generate personalised diagnostic reports for both clinicians and patients.

**Chapter 6** discusses the key findings of the research and outlines potential directions for future work.

## CHAPTER 2: LITERATURE SURVEY

### **2.1 Introduction**

In recent years, the integration of AI into medical diagnostics has significantly transformed research and clinical practices related to neurodegenerative disorders, particularly PD. Among the various AI techniques, ML and DL have shown immense potential in identifying disease-specific biomarkers, classifying disease states, and supporting early-stage diagnosis. These techniques have been effectively applied across diverse data types, including neuroimaging, electrophysiological signals, clinical assessments, and multimodal biomarker information. As AI in medicine continues to evolve, increasing attention has shifted toward the interpretability of models, especially in high-stakes domains like neurology. The emergence of hybrid architectures combining multiple data types and optimization strategies reflects a growing need for transparent and clinically interpretable decision-making tools. In this context, DL models particularly CNNs, recurrent networks (LSTMs), and attention-based systems have played a pivotal role by automating feature extraction and enabling multimodal fusion.

More recently, the advent of Large Language Models (LLMs) such as ChatGPT and BioGPT has expanded the boundaries of medical AI. These models offer powerful capabilities for integrating unstructured text (e.g., clinical notes, radiology reports) with structured visual or numeric data, enabling tasks such as clinical report generation, decision support, and question answering. Despite these advancements, several challenges persist. Limitations such as dataset imbalances, difficulties in generalizing findings across diverse clinical contexts, and the complexities involved in effectively integrating heterogeneous data modalities underscore the need for further refinement and innovation in AI-based diagnostic approaches.

To address these gaps, this chapter systematically explores ML, DL, and LLM-based methods individually, clearly highlighting their evolution, distinct strengths, and inherent limitations.

Specifically, Section 2.2 examines traditional neuroimaging techniques, summarizing their effectiveness and current diagnostic performance while highlighting modality-specific challenges. Section 2.3 delves into ML-driven classification strategies that utilize multimodal data, illustrating both their advantages and constraints related to feature engineering and modality integration. Subsequently, Section 2.4 critically evaluates state-of-the-art DL methodologies, emphasizing their automated feature extraction capabilities, enhanced diagnostic accuracy, and the challenges regarding interpretability and clinical validation. Finally, Section 2.5 explores the transformative role of LLMs, focusing on their unique potential for multimodal reasoning and their current limitations within clinical applications. The insights gained from this comprehensive literature review emphasize the significance of integrating multimodal data to overcome existing diagnostic limitations. Ultimately, this analysis lays a robust foundation for the integrated multimodal diagnostic frameworks proposed in subsequent chapters, which aim to achieve improved diagnostic precision, clinical relevance, and interpretability for effective management and monitoring of PD.

## **2.2 PD diagnosis using Neuroimaging data**

Recent advancements in computational neuroimaging have substantially enhanced diagnostic capabilities in the context of PD. Neuroimaging modalities, such as MRI, functional MRI (fMRI), Single Photon Emission Computed Tomography (SPECT), Positron Emission Tomography (PET), and Dopamine Transporter scans (DaTSCAN), have been extensively employed due to their ability to detect subtle yet clinically significant structural and functional brain alterations associated with PD.

### **2.2.1 Structural MRI and Morphometric Analysis**

MRI, in particular, has been widely adopted for assessing structural abnormalities, primarily in regions like the substantia nigra and associated cortical areas. Marino et al. (2012) utilized advanced MRI techniques, including magnetic resonance spectroscopy (MRS), diffusion-weighted imaging (DWI), diffusion tensor imaging (DTI), and fMRI, highlighting

abnormalities within the olfactory system indicative of the prodromal stage of PD, characterized by significant dopaminergic neuron loss [16]. Voxel-based morphometry (VBM) approaches have gained prominence for detecting cortical atrophy. Solana-Lavalle et al. (2021) and Zhai et al. (2023), reported diagnostic accuracies exceeding 96% and identified extensive cortical degeneration associated with PD progression [33] and [34]. Additionally, Yang et al. (2020) proposed the superior temporal gyrus as a potential biomarker for PD-Rapid Eye Movement Sleep Behavior Disorder (PD-RBD) subtypes based on significant grey matter reductions [35]. Chakraborty et al. (2020) employed a 3D feature extraction pipeline using Artificial Neural Networks (ANN), achieving a diagnostic accuracy of 98.51% by analyzing textural and morphometric variations in substantia nigra [36].

### **2.2.2 Iron Deposition and Quantitative MRI Biomarkers**

Tambasco et al. (2019) investigated iron accumulation using T2\*-weighted MRI, correlating it with motor and cognitive dysfunctions. Although no accuracy metrics were reported, the study emphasized the potential of iron deposition as a progression marker [15]. Zeng et al. (2021) reviewed neurochemical biomarkers measurable via high-field MRS, identifying alterations in the NAA/Cr and MI/Cr ratios in the hippocampus and entorhinal cortex—findings relevant to prodromal PD despite being based on Alzheimer's datasets [12].

### **2.2.3 Functional and Molecular Imaging with AI Models**

Functional and molecular imaging techniques, including PET and SPECT, have demonstrated exceptional sensitivity in detecting presymptomatic dopaminergic deficits. Notably, Pahuja et al. (2020) integrated DaTSCAN SPECT imaging with biological biomarkers using deep learning and multivariate logistic regression, achieving remarkable diagnostic accuracy of 100%, highlighting the potential of multimodal integration for early PD detection [37]. Similarly, Yoon et al. (2021) achieved a diagnostic accuracy of 95.7% by analyzing texture features from PET imaging using advanced ML algorithms like XGBoost [38]. Furthermore, the combination of CNN and explainable AI (XAI) techniques, as shown by Pianpanit et al.

(2021), further emphasized the importance of model interpretability and clinical trust, with accuracies surpassing 90% [39]. Studies employing resting-state fMRI, such as those by Shi et al. (2022), reinforced the significance of functional connectivity measures in identifying early-stage PD, albeit with moderate accuracy levels around 78% [40].

Mortezaee et al. (2021) highlighted the high sensitivity (87.7%) of DaTSCAN SPECT and the efficacy of PET imaging in detecting presymptomatic dopaminergic degeneration, affirming their utility in differential diagnosis and early-stage PD identification [41]. Abbasi et al. (2024) addressed a critical challenge in multi-center neuroimaging studies—scanner-induced variability. Their review emphasized the use of generative adversarial networks (GANs) and Transformers for harmonizing structural MRI data, which is essential for achieving reproducible and generalizable diagnostic models [42]. Table 2.1 summarizes key studies employing neuroimaging modalities and associated AI methods, clearly outlining achieved diagnostic accuracies, employed methodologies, and clinical implications.

#### **2.2.4 Synthesis and Transition to ML Section**

Despite significant diagnostic accuracies achieved using various neuroimaging modalities, persistent challenges limit their broader clinical applicability. Key issues include scanner-induced variability, constrained generalizability across diverse datasets, dependency on manual interpretation, and the complexity inherent to extracting robust, discriminative features from high-dimensional imaging data. These modality-specific constraints necessitate a shift towards more sophisticated computational approaches. Consequently, recent research has increasingly leveraged advanced ML algorithms, particularly those capable of automated feature extraction and improved diagnostic consistency, as comprehensively explored in the next section.

**Table 2.1** Summarises the diagnosis of PD based on neuroimaging techniques using AI approaches

<b>Authors &amp; References</b>	<b>Modality Used</b>	<b>No. of Subjects</b>	<b>Methodology</b>	<b>Results and evaluation</b>
[16] Marino et al. (2012)	MRI, MRS, DWI, DTI, fMRI	Not specified	DTI, fMRI analysis	Prodromal stage detection
[33] Solana-Lavalle et al. (2021)	MRI	Gender-specific dataset	VBM	Males: 99.01%, Females: 96.97%
[34] Zhai et al. (2023)	MRI	87 PD, 47 HC	VBM	Extensive cortical degeneration observed
[43] Yang et al. (2020)	MRI	105 PD+RBD, 140 PD–RBD	VBM meta-analysis	Superior temporal gyrus as biomarker
[36] Chakraborty et al. (2020)	MRI (PPMI dataset)	Not specified	3D ANN	Accuracy: 98.51%
[12] Zeng et al. (2021)	MRI, PET, SPECT, MRS (AD datasets)	Not applicable	Review of neurochemical markers	Highlighted decreased NAA/Cr and increased MI/Cr ratios in hippocampus and cingulate gyrus
[37] Pahuja et al. (2020)	DaTSCAN SPECT	384 PD, 148 HC	Stacked autoencoder, MLR	Accuracy: 100%
[39] Pianpanit et al. (2021)	SPECT	Not specified	CNN, XAI	Accuracy: >90%
[15] Tambasco et al. (2019)	T2*-MRI	32 PD	Iron content correlation	Biomarker for progression

Table 2.1 (continued...)

[38] Yoon et al. (2021)	PET	31 PD, 31 HC	XGBoost	Accuracy: 95.7%
[40] Shi et al. (2022)	rs-fMRI	123 PD, 90 HC	ALFF radiomics, SVM	Accuracy: 78.07%, Specificity: 76.08%, Sensitivity: 78.80%
[42] Abbasi et al. (2024)	Structural MRI	Not specified	GANs, Transformers	Focus on scanner harmonization in multi-site studies

### 2.3 ML models for PD classification using Multimodality data

ML has brought remarkable changes to healthcare, particularly in enhancing the diagnosis and ongoing monitoring of neurodegenerative diseases like PD. The integration of multimodal data comprising neuroimaging features, biological biomarkers, vocal patterns, cerebrospinal fluid (CSF) characteristics, and clinical assessments has markedly enhanced the accuracy and robustness of early-stage PD detection. Recent studies underscore the increasing significance of multimodal analytics for achieving superior diagnostic outcomes. Zhang et al. (2022) reviewed ML methodologies applied to neuroimaging techniques (SPECT, PET, structural MRI, resting-state fMRI), highlighting the effectiveness of Support Vector Machines (SVM), Artificial Neural Networks (ANN), CNNs, and Random Forest classifiers in achieving accuracies between 95% and 98% across diverse datasets [20]. For instance, integrating patient age with radiotracer uptake demonstrated distinct advantages over traditional semi-quantitative methods, particularly for prodromal stage identification.

Ya et al. (2022) provided further validation of multimodal ML approaches, utilizing structural MRI data from multiple brain regions. By employing feature selection through Least Absolute Shrinkage and Selection Operator (LASSO) and classification via logistic regression, their multimodal model significantly outperformed single-region methods, achieving superior Area

Under Curve (AUC) scores in both training (0.781) and testing (0.756) datasets [44]. Critical predictors such as cortical thickness, gyration indices, and asymmetry reinforced the importance of integrating multiple anatomical features. Beyond neuroimaging, studies exploring voice and speech biomarkers have underscored their potential as effective non-invasive tools. Thakur et al. (2022) evaluated multiple classifiers (e.g., Extra Trees, Random Forests, SVM) on voice samples, achieving highest accuracy (93.39%) with Extra Trees, indicating that vocal biomarkers can reliably differentiate PD patients from healthy individuals [45]. Similarly, Rana et al. (2022) achieved an accuracy of 87.17% using Artificial Neural Networks (ANN) on acoustic features, highlighting speech analysis as a complementary diagnostic modality [46].

Cerebrospinal fluid (CSF) biomarkers offer high diagnostic specificity due to their proximity to central nervous system pathology. George et al. (2025) analyzed CSF proteomics using omics-based clustering and dimensionality reduction, uncovering subtype-specific markers that distinguish Multiple System Atrophy (MSA) from PD. Tumani et al. (2019) outlined clinical guidelines for CSF interpretation in neurodegenerative conditions, emphasizing its utility in differential diagnosis. However, CSF data remains limited due to the invasive nature of collection and the small sample sizes available for ML training. Furthermore, Lotankar et al. (2017) surveyed biochemical disruptions in CSF, including alpha-synuclein, total tau, and phosphorylated tau, all of which showed consistent reductions in PD patients. The study recommended multifactorial biomarker integration to improve early-stage monitoring accuracy.

Siderowf et al. (2023), conducted a landmark study across 1,123 participants, applying alpha-synuclein seed amplification assays (SAA) to CSF samples. Their ML-enhanced diagnostic model achieved 88% sensitivity and 96.3% specificity. However, performance was significantly lower in LRRK2 mutation carriers, indicating the biological heterogeneity of PD subtypes and highlighting the need for subtype-aware models [47]–[50]. Furthermore, CSF biomarkers such as beta-amyloid<sub>1–42</sub>, tau, and phosphorylated-tau (p-tau) have consistently demonstrated correlations with cognitive and motor deterioration in PD, reinforcing their

clinical utility [10]. Table 2.2 summarizes key ML-based studies for PD classification using multimodal data, clearly outlining methodologies, and main findings.

### 2.3.1 Synthesis and Transition to DL Section

While the reviewed ML methods demonstrate strong diagnostic performance, their dependence on handcrafted features, sensitivity to data imbalance, and restricted capacity for deep multimodal integration significantly limit their clinical scalability and robustness. These constraints underscore the necessity for methodologies capable of automated and sophisticated feature extraction, enhanced generalizability, and effective multimodal fusion. Consequently, the progression towards DL methods emerges logically, providing powerful automated feature extraction capabilities and superior scalability, comprehensively addressed in the following section.

**Table 2.2** Summary of ML Methods for PD Detection (Multimodal)

Author & reference	Dataset	Feature selection	Methods used	Main findings	Other findings
[20] Zhang et al. (2022)	Neuroimaging (SPECT, PET, MRI, fMRI)	Not explicitly detailed (review)	SVM, ANN, CNN, RF	Machine learning significantly improves PD detection vs traditional methods	Multimodal imaging with ML enhances diagnostic accuracy
[44] Yang et al. (2022)	Own: 60 PD, 56 HC; PPMI: 69 PD, 71 HC	LASSO, Pearson correlation	Logistic Regression	Combined model AUC = 0.781 (train), 0.756 (test)	Cortical thickness, gyration & asymmetry indices were key features

Table 2.2 (Continued).....

[46] Rana et al. (2022)	UCI PD Dataset (Voice)	Not specified	SVM, KNN	SVM achieved 91.3% accuracy	Voice analysis is effective for early non-invasive PD detection
[47] George et al. (2025)	CSF samples (Proteome & Metabolome )	Omics-based clustering, dimensionality reduction	Bioinformatics + ML-based clustering	Identified biomarkers distinguishing MSA vs PD	Multi-omics approaches provide deep insights into neurodegenerative mechanisms
[48] Tumani et al. (2020)	CSF Clinical Guidelines	N/A	Clinical CSF analysis	CSF analysis is critical in neurodegenerative differential diagnosis	Guidelines highlight clinical best practices for lumbar puncture and CSF interpretation
[49] Lotanka r et al. (2017)	CSF-based biomarker literature	Multiple (e.g., alpha-beta synuclein, t-tau, p-tau)	Review of biochemical/clinical methods	CSF biomarkers critical in early PD detection	Multifactorial biomarker integration needed for early-stage monitoring
[50] Siderow f et al. (2023)	PPMI, international CSF samples	Alpha synuclein seed amplification assay	Assay-based sensitivity/specificity tests	88% sensitivity, 96.3% specificity	Reduced sensitivity in LRRK2 variant carriers suggests biological PD subtype variation

## 2.4 DL models for PD classification using Multimodality data

Advancements in neuroimaging technologies, when combined with DL methods, have significantly enhanced the diagnostic accuracy, interpretability, and clinical relevance of PD classification systems. In contrast to traditional ML approaches, DL frameworks—such as CNNs, Long Short-Term Memory (LSTM) networks, Graph Neural Networks (GNNs), and hybrid explainable AI (XAI) models enable automatic feature extraction from high-dimensional multimodal datasets. This section categorizes DL-based PD studies by the primary data modality used: electrophysiological (EEG), neuroimaging (MRI, SPECT, DTI), clinical and cerebrospinal fluid (CSF) biomarkers, and multimodal integration strategies.

### 2.4.1 EEG-Based Deep Learning Approaches

Electroencephalography (EEG), due to its non-invasive nature and high temporal resolution, has been effectively used in DL frameworks for PD classification. Obayya et al. (2023) proposed a Densely Linked Bidirectional LSTM (DLBLSTM) model applied to EEG spectrograms, achieving a remarkable accuracy of 99.6% using six-fold cross-validation. However, the dataset size was limited, and external validation was not conducted, raising concerns regarding generalizability [23]. Nour et al. (2023) extended EEG-based classification by introducing a 1D Partial Distance Covariance Neural Network (1D-PDCovNN), coupled with Independent Component Analysis (ICA) and Common Spatial Pattern (CSP) filtering. Their approach achieved 99.31% accuracy and leveraged a dynamic classifier selection mechanism based on a meta-learning algorithm (DCS-MLA), thus improving robustness across noisy EEG environments [51].

### 2.4.2 Neuroimaging and Explainable CNN Frameworks

DL models using neuroimaging modalities such as MRI, SPECT, and PET have been instrumental in detecting structural and functional brain changes associated with PD. Magesh et al. (2020) applied transfer learning using VGG16 on DaTSCAN SPECT images and utilized

Local Interpretable Model-agnostic Explanations (LIME) to visualize predictive regions. The model achieved 95.2% accuracy, with LIME highlighting the putamen and caudate as significant features [52]. Similarly, Nazari et al. (2022) trained a CNN on DaT-SPECT data and implemented Layer-wise Relevance Propagation (LRP) to identify critical regions, achieving 95.8% accuracy, 92.8% sensitivity, and 98.7% specificity [53]. Camacho et al. (2023) trained a CNN on 2,041 T1-weighted MRI scans and incorporated Jacobian determinant maps with saliency-based interpretation. The model achieved an accuracy of 79.3% and an AUC-ROC of 0.87, identifying the frontal and deep gray matter regions as relevant biomarkers [54]. Yang et al. (2023) implemented a 3D ResNet model on whole-brain MRI for PD classification, achieving 96.1% cross-validated accuracy. Grad-CAM visualizations revealed that activations in the frontal lobe correlated with disease severity based on Unified Parkinson's Disease Rating Scale (UPDRS) scores [55].

#### **2.4.3 Clinical, CSF Biomarker, and Graph-Based Deep Learning**

Wang et al. (2020) designed a multimodal DL model incorporating non-imaging features such as REM sleep behavior, olfactory deficits, CSF biomarkers, and DaTSCAN indicators. The model achieved 96.45% accuracy and provided feature importance rankings, enabling premotor screening [56]. Nilashi et al. (2023) developed a hybrid model combining Deep Belief Networks (DBNs) with Neuro-Fuzzy inference systems to predict UPDRS scores. Their architecture, enhanced through Principal Component Analysis (PCA) and Expectation-Maximization (EM) clustering, improved real-time monitoring accuracy while reducing model complexity [57]. Huang et al. (2023) introduced MNC-Net, a Graph Neural Network (GNN) that used structural brain network clustering from DTI data. The model achieved 95.5% accuracy and highlighted brain regions associated with mood, memory, and motor functions using saliency maps [58].

#### **2.4.4 Multimodal Fusion Strategies**

Islam et al. (2024) conducted a meta-analysis of 60 studies applying DL to handwriting and voice datasets. Ensemble models were found to consistently outperform single-modality approaches, achieving accuracy rates up to 98.5%. The review emphasized the need for standardized benchmarking across heterogeneous datasets [21]. Pahuja and Prasad (2022) compared modality-level versus feature-level fusion strategies for MRI, SPECT, and CSF. Their results showed that feature-level fusion yielded superior performance (93.33% accuracy), supporting the case for deeper integration of heterogeneous data sources [59].

#### **2.4.5 Hybrid and Transferable DL Architectures**

Several studies have explored cross-domain deep learning models originally designed for other neurological disorders and demonstrated their adaptability to PD. Balasubramanian et al. (2024) applied a hybrid model integrating Residual Networks with Shepherd CNN (RF-ShCNN) for PD-related MRI analysis, achieving 94% accuracy. Though originally developed for brain tumor classification, this architecture proved flexible for neurodegenerative applications [60]. Mohammad and Al Ahmadi (2023) used VGG19 feature extraction along with the Whale Optimization Algorithm (WOA) for Alzheimer's diagnosis, reaching 99% accuracy. This work illustrates the utility of evolutionary feature selection in enhancing CNN-based classification performance, with techniques applicable to PD research [61].

Cheng et al. (2024) proposed a residual attention fusion model with decorrelation strategies for Alzheimer's and Mild Cognitive Impairment (MCI) detection, achieving 86.79% accuracy. While AD-focused, the model's robust fusion design shows strong potential for adaptation to multimodal PD classification [62]. Salmanpour et al. (2021) explored multimodal classification and PD progression forecasting through motor, non-motor, and radiomics features from SPECT images. Their hybrid ML framework exceeded 90% accuracy, showcasing radiomics' vital role in distinguishing PD subtypes and progression trajectories [19]. Hamza et al. (2022) initially designed deep Bayesian-optimized frameworks (D2BOF-COVIDNet) for COVID-19,

achieving near-perfect accuracies (up to 99.9%). These transferable models demonstrated strong potential for PD diagnostics, notably through interpretative Grad-CAM visualizations [63], [64]. DL methodologies inadequately handle unstructured clinical narratives, limiting their potential to leverage rich textual patient histories and clinical documentation. Table 2.3 succinctly summarizes pivotal studies utilizing DL techniques, clearly outlining methodologies, performance metrics, and interpretability mechanisms.

#### 2.4.6 Synthesis and Implications

Deep learning models have substantially advanced PD diagnosis through high classification accuracy, automatic feature extraction, and multimodal fusion capabilities. Nevertheless, challenges remain: (i) Interpretability tools like LIME, Grad-CAM, and LRP help, more transparent integration with clinical variables is needed. (ii) Many models use single-center or augmented datasets with limited external validation. (iii) Imbalanced classes and improperly aligned multimodal data can reduce diagnostic robustness. These gaps motivate the design of explainable, radiomics-driven DL frameworks that can integrate neuroimaging with clinical and molecular data. The next section explores the growing role of Large Language Models (LLMs), which enable reasoning over unstructured data and open new avenues for multimodal AI in healthcare.

**Table 2.3** Summary of DL Methods for PD Detection (Multimodal)

Author & Year	Dataset	Feature Selection	Methods Used	Main Findings	Other Findings
[23] Obayya et al. (2023)	EEG Spectrogram	EEG signal pre-processing	DLBLSTM (Bidirectional LSTM)	Achieved 99.6% accuracy using 6-fold cross-validation	EEG is a promising non-invasive modality for PD diagnosis
[21] Islam et al. (2024)	Handwriting & Voice	Review-based (not specified)	CNN, LSTM,	Ensemble DL reached up to 98.5% accuracy	Multimodal data outperforms single-modal methods

Table 2.3 (Continued)

	(review of 60 studies)		Ensemble DL		
[52] Magesh et al. (2020)	DaTSCAN SPECT (PPMI)	Transfer learning (VGG16)	CNN + LIME	95.2% accuracy, 97.5% sensitivity, 90.9% specificity	LIME identifies critical regions like putamen and caudate
[54] Camacho et al. (2023)	T1-weighted MRI (2041 subjects, PPMI)	Jacobian maps + clinical features	CNN + Saliency maps	79.3% accuracy, AUC-ROC 0.87	Frontal and deep gray matter regions were key
[53] Nazari et al. (2022)	DAT-SPECT (1296 total, PPMI)	CNN learned features	CNN + LRP	Accuracy 95.8%, Sensitivity 92.8%, Specificity 98.7%	LRP enhanced interpretability by focusing on putamen
[56] Wang et al. (2020)	REM, CSF, Olfaction, DaTSCAN	Premotor feature fusion	DL + Boosting	Accuracy 96.45%	Enabled premotor screening with ranked features
[57] Nilashi et al. (2023)	Real-world UPDRS dataset	PCA, EM Clustering	DBN + Neuro-Fuzzy	Improved UPDRS prediction accuracy	Reduced complexity for real-time PD monitoring
[51] Nour et al. (2023)	EEG (Resting-state)	ICA + CSP	1D-PDCovN N + DCS-MLA	99.31% accuracy in EEG-based PD classification	Robust against noise; high interpretability with multiple classifiers
[58] Huang et al. (2023)	DTI-based Structural Brain Networks	Graph node clustering	MNC-Net (GNN)	95.5% accuracy	Saliency maps revealed mood/memory/motor-linked regions
[62] Cheng et al. (2024)	MRI + FDG-PET (ADNI)	Decorrelation + Mutual Attention Fusion	Residual Attention Fusion DL	86.79% accuracy for AD/MCI/NC	Technique adaptable to PD multimodal diagnosis
[55] Yang et al. (2023)	T1-weighted MRI (PPMI)	Voxel-based intensity	3D ResNet + Grad-CAM	96.1% (cross-val), 94.5% (hold-out)	Frontal lobe critical in early PD with semantic correlation

Table 2.3 (Continued)

[61] Mohamma d & Al Ahmadi (2023)	MRI (AD- focused)	Whale Optimizatio n Algorithm (WOA)	VGG19 + F-KNN	Accuracy 99%, Precision 99%	Feature selection greatly enhanced diagnostic performance
---	----------------------	--	------------------	--------------------------------	--

## 2.5 Large Language Models (LLMs) in medical field

Large Language Models (LLMs) have substantially transformed the scope of artificial intelligence within the medical field, providing capabilities far exceeding those of traditional ML and DL approaches. Unlike conventional models, which primarily utilize structured datasets, LLMs excel in processing extensive volumes of unstructured textual and multimodal clinical information. This capability enables advanced applications such as automated clinical report generation, multimodal diagnostic reasoning, patient-specific analysis, and interactive clinical decision support systems.

Recent studies highlight the significant potential and current limitations of LLMs within various clinical domains. For example, El Haj et al. (2023) evaluated ChatGPT's ability to interpret neuropsychological assessments for Alzheimer's Disease, identifying gaps in nuanced clinical reasoning, particularly regarding specialized cognitive deficits [65]. Similarly, Yang et al. (2024) observed marked improvements in ChatGPT's diagnostic accuracy for bone tumors following domain-specific fine-tuning, underscoring the critical role of targeted adaptation for medical applications [66]. Malik and Zaheer (2024) explored the integration of ChatGPT into cancer pathology workflows, demonstrating the model's effectiveness in structured data extraction from pathology reports, although emphasizing ongoing concerns regarding clinical validation, regulatory compliance, and ethical deployment [67].

Additionally, Bartoli et al. (2024) and Campbell et al. (2024) highlighted both strengths and shortcomings of LLMs in clinical education and radiology, particularly noting their utility in multilingual patient communication and clinical workflow support, alongside recognized issues

such as 'hallucinations' and inconsistent outputs in complex clinical contexts [68] and [69]. In more specialized applications, Zhu et al. (2024) successfully utilized ChatGPT for generating structured exercise prescriptions, but highlighted limitations regarding nuanced clinical judgment, reinforcing the supplementary rather than autonomous clinical role of current LLMs [70]. Liu et al. (2023) directly compared GPT-3.5 and GPT-4.0 with experienced neurosurgeons, revealing GPT-4.0's superior capacity for complex clinical reasoning, further supporting the progressive potential of domain-adapted LLMs for high-level medical decision-making [71].

Karakas et al. (2023) investigated practical LLM integration within pediatric neurology clinics, finding ChatGPT notably effective for administrative tasks like generating medical necessity documentation and educational materials. Despite enhancing workflow efficiency, its application remains limited for complex clinical decision-making, highlighting the essential role of clinician oversight. In radiology, Tippareddy et al. (2023) explored LLMs for radiology education, documentation automation, and burnout reduction. While these tools enhanced productivity and educational value, the risk of factual errors persisted [72], [73]. Hu et al. (2024) further validated ChatGPT's competitiveness with traditional ML models in radiology report interpretation for lung cancer. However, they noted that over-reliance on prompt engineering could create inconsistencies across tasks [74]. However, Shifai et al. (2024), evaluating ChatGPT Vision's performance in melanoma classification, cautioned against standalone usage due to limited diagnostic accuracy, further underscoring the need for hybrid approaches [75]. While demonstrating considerable promise, the application of LLMs in medical diagnosis, including PD, is constrained by several limitations. Foremost among these are inconsistent clinical reasoning, susceptibility to "hallucinations" (generation of inaccurate or misleading information), and inadequate domain-specific adaptation, which restrict their clinical reliability. Furthermore, the validation of LLM-based clinical tools remains insufficient, raising legitimate concerns about their safe and effective deployment in clinical environments. Additionally, their

capability for multimodal integration and reasoning, crucial in PD diagnosis, requires further development and validation.

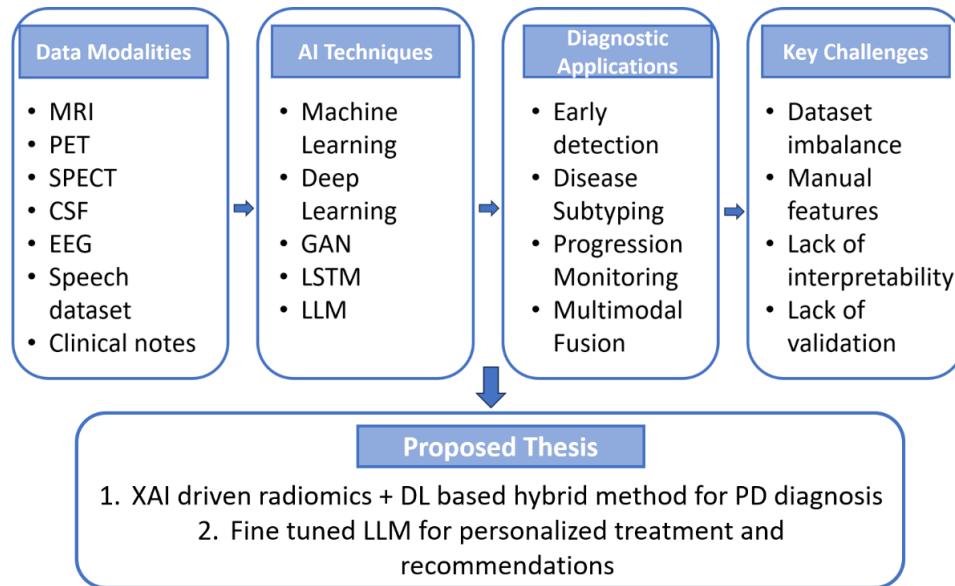
### **2.5.1 Synthesis and Explicit Linkage to Research Goals**

Large Language Models (LLMs) demonstrate significant promise due to their ability to integrate and interpret unstructured textual and multimodal clinical data. However, critical limitations such as inconsistent clinical reasoning, hallucination of facts, limited multimodal reasoning, and insufficient domain-specific fine-tuning highlight the need for robust, validated, and hybrid diagnostic frameworks. These identified gaps directly motivate the methodological innovations proposed in this thesis. Specifically, the proposed approach strategically combines the strengths of DL, particularly automated feature extraction and modality-specific accuracy, with the advanced reasoning capabilities of fine-tuned LLMs. This integrated framework aims to deliver accurate, robust, clinically interpretable, and externally validated diagnostic predictions for PD, which are further detailed in the chapters that follow. To visually synthesize the methodological trajectory covered in this chapter. Figure 2.1 presents a conceptual overview of the evolution of AI-based PD diagnosis. It illustrates the progression from various clinical data modalities to AI techniques, their diagnostic applications, and key challenges, culminating in the proposed thesis framework that unifies XAI-driven radiomics with fine-tuned LLMs for early diagnosis and personalized treatment.

### **2.6 Summary**

This chapter has presented an extensive literature review examining recent advancements and methodologies in AI for the diagnosis and classification of PD. It systematically explored the applications, advantages, and limitations of traditional neuroimaging approaches, ML, DL, and Large Language Models (LLMs), emphasizing their roles across diverse multimodal clinical datasets. Despite high diagnostic accuracies, neuroimaging-based methods remain limited by

scanner-induced variability, modality-specific constraints, and dependence on manual interpretation. ML approaches improve flexibility through multimodal fusion and handcrafted features, but often lack scalability and robustness across diverse datasets. DL models enable automated feature learning and multimodal integration, but are frequently hindered by interpretability issues, lack of external validation, and challenges in handling unstructured clinical narratives. LLMs represent a transformative leap in clinical reasoning and multimodal understanding, yet their domain adaptation limitations and susceptibility to hallucinations restrict current real-world adoption.



**Figure 2.1:** Evolution of AI methods in PD diagnosis, highlighting data modalities, techniques, clinical applications, key challenges, and the proposed thesis framework.

The key unresolved gaps identified across the literature include:

- Limited performance in prodromal or early-stage PD diagnosis.
- Inadequate multimodal fusion of imaging, clinical, and biochemical data.

- Lack of model interpretability and clinical trust.
- Insufficient external validation and generalization across real-world settings.

These gaps highlight the pressing need for an integrated, explainable, and clinically viable diagnostic framework one that can fuse radiomic features from imaging data with structured clinical and molecular information, while leveraging the reasoning capabilities of advanced LLMs. Building upon these insights, the following chapter introduces a novel XAI-driven radiomics framework for early PD detection using subcortical MRI data, enriched with fine-tuned language models to support clinical interpretation and multimodal reasoning.

## CHAPTER 3: DIAGNOSIS OF PD BASED ON MRI SCANS

### 3.1 Introduction

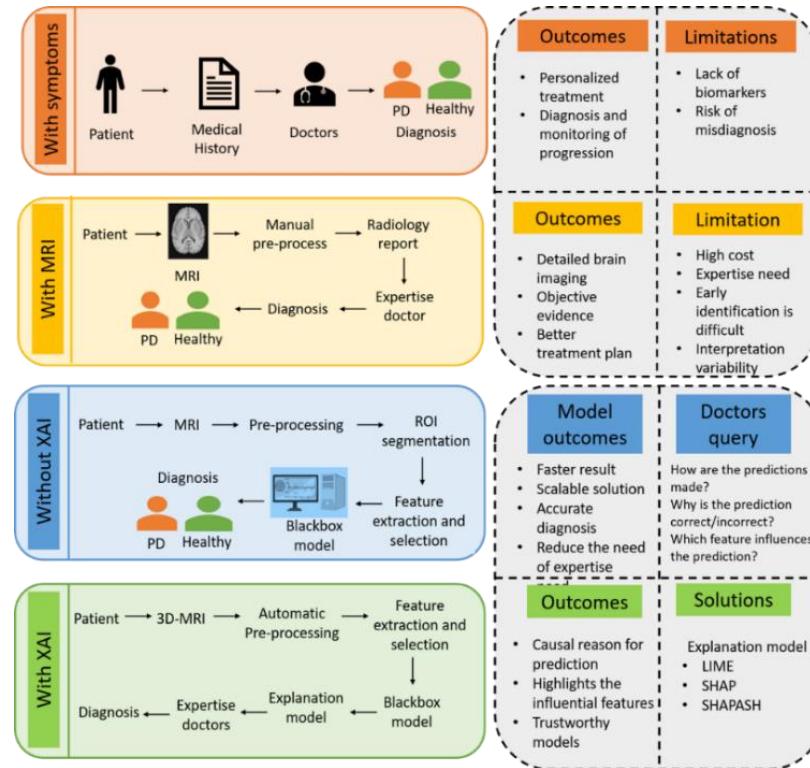
PD is a long-term neurological condition that gradually affects movement. People with PD often experience symptoms such as shaking, muscle stiffness, slowed physical movements, and difficulty with balance and coordination. These symptoms stem from the gradual degeneration of dopamine-producing neurons in the SN of the brain. Beyond motor impairments, individuals with PD frequently experience a range of non-motor symptoms such as cognitive decline, depression, anxiety, and sleep disturbances, all of which significantly diminish overall quality of life [76]–[78]. Although early detection of PD has the potential to substantially improve clinical outcomes by enabling timely therapeutic interventions, current diagnostic practices remain predominantly based on clinical symptom assessment. Unfortunately, such methods often fail to identify PD at an early stage, as symptoms typically manifest only after considerable neuronal loss and pronounced pathological changes such as iron accumulation—have occurred in subcortical brain regions [6], [7]. In the absence of a definitive cure, the development of accurate and early-stage diagnostic techniques remains crucial for optimising therapeutic strategies and improving patient management.

In recent years, breakthroughs in neuroimaging technologies like MRI, PET, and SPECT have greatly improved our ability to detect PD at earlier stages, offering new possibilities for timely diagnosis and intervention [12], [41]. Among these, MRI has emerged as a particularly valuable diagnostic tool owing to its high-resolution capabilities, non-invasive nature, and the absence of ionizing radiation. These attributes collectively enable safe and detailed visualization of structural brain alterations associated with PD progression. Among various MRI techniques, T2-weighted sequences have shown greater sensitivity in detecting subtle changes linked to PD, such as iron buildup in the brain. In contrast, sequences like T1-weighted and FLAIR are

generally less effective at highlighting these specific biomarkers.[15], [36]. Despite these advantages, clinical adoption of MRI remains constrained due to the subtle nature of early-stage biomarkers, high inter-observer variability in interpretation, and associated costs, thereby limiting its widespread and routine use. To address these challenges, emerging computational methodologies based on ML and DL have gained increasing attention. These techniques can automatically detect complex and nuanced neuroimaging patterns associated with early-stage PD, thereby reducing dependence on subjective clinical interpretation and significantly improving diagnostic accuracy [21], [23], [79]. However, the inherent complexity of ML and DL models often renders them as 'black-box' systems, lacking interpretability and transparency in their decision-making processes. This lack of explainability presents a substantial barrier to clinical integration, as healthcare professionals require a clear understanding of model-driven insights to support evidence-based clinical decision-making.

To address the critical challenge of model interpretability, this research integrates Explainable Artificial Intelligence (XAI) methodologies with robust ML techniques, aiming to enhance clinical transparency and support informed decision-making. By employing XAI tools such as Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), the study provides clear insights into how specific MRI-derived radiomic features influence diagnostic outcomes. This interpretability not only bolsters clinician confidence but also supports the development of personalized patient management strategies. This chapter specifically explores a wide array of radiomic features, including structural, morphological, textural, and statistical attributes extracted from MRI scans of subcortical brain areas that are heavily impacted by PD. The complete methodological framework integrating MRI preprocessing, radiomic feature extraction, ML-based classification, and XAI interpretation is systematically presented and illustrated in Figure 3.1. In summary, the key contributions of this chapter include: (i) the development of an advanced diagnostic framework that combines MRI imaging, ML, and XAI techniques to identify early-stage PD accurately; (ii) the enhancement of model interpretability to promote clinical trust and adoption; and (iii) the improvement of

predictive performance through careful feature selection and optimized data handling. Ultimately, this work aspires to equip clinicians with an interpretable and effective tool for timely PD detection and personalized therapeutic planning.

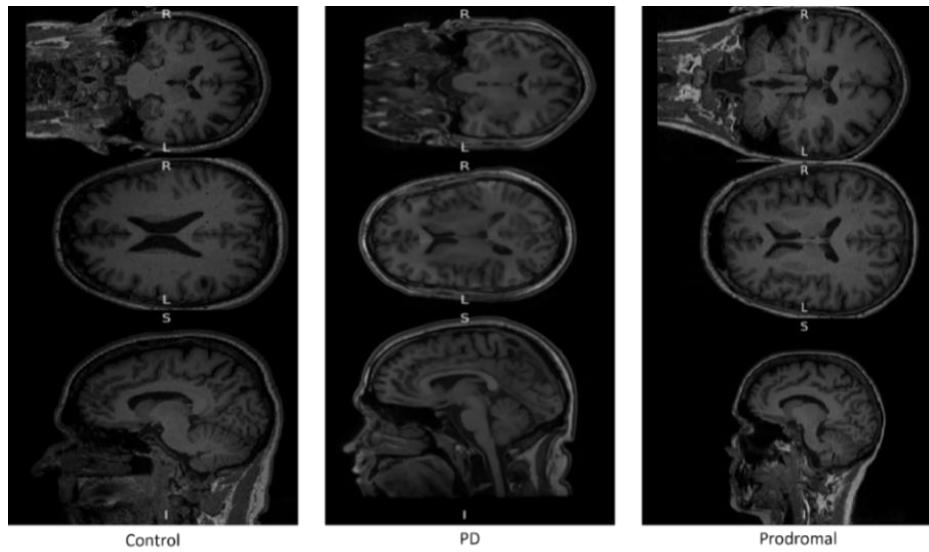


**Figure 3.1** Workflow of the designed system integrating XAI and ML approaches for early diagnosis and progression tracking of PD

### 3.2 Dataset description

In this study, the MRI data used for analysis were sourced from the Parkinson's Progression Markers Initiative (PPMI) database ([www.ppmi-info.org](http://www.ppmi-info.org)), a globally recognized, multicenter collaborative project designed to identify and validate biomarkers for PD. The PPMI dataset is particularly well-suited for this research due to its large sample size, standardized imaging protocols, and public availability, which promote both reproducibility and comparative

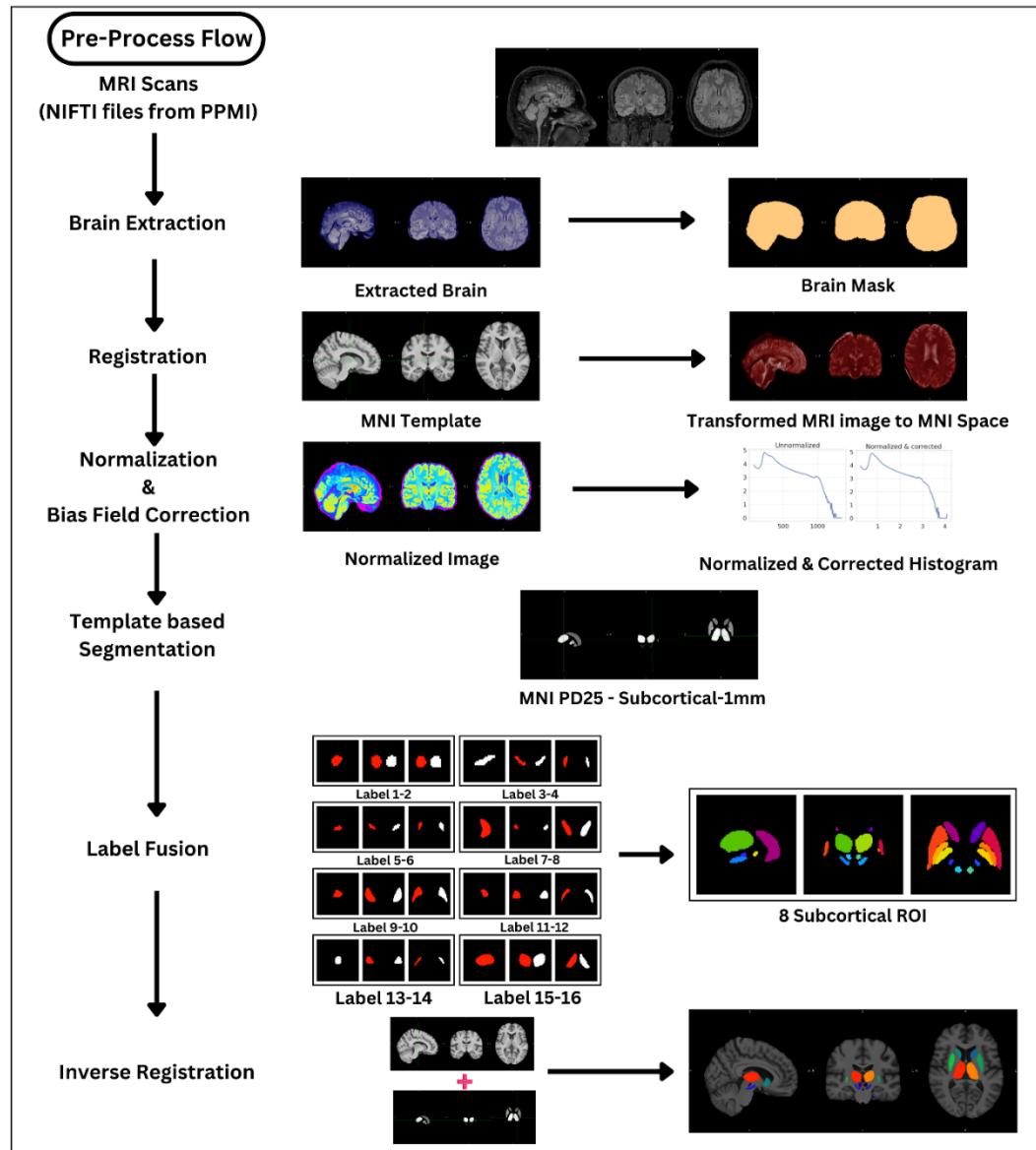
benchmarking across PD studies. The dataset includes a total of 500 participants, divided into three key clinical groups: 180 diagnosed with PD, 160 healthy control (HC) subjects, and 160 prodromal individuals who may be at risk of developing PD. The average ages within the groups are relatively consistent: 61.9 years for PD patients, 59.9 years for healthy controls, and 60.2 years for prodromal cases, with balanced gender distribution across all categories. Each participant underwent baseline MRI scanning using a 3.0 Tesla Siemens MRI scanner following a 3D sagittal FLAIR sequence protocol. This protocol was chosen based on its high contrast resolution and compatibility across imaging centres. The scans covered essential anatomical regions implicated in PD, including the pons, cerebellum, and vertex. The MRI acquisition parameters included a Spin Echo (SE)/Inversion Recovery (IR) pulse sequence with T2-weighting, allowing for enhanced visualization of tissue contrast. Scans were acquired with 1.0 mm slice thickness, matrix dimensions between 256×256 and 512×512 pixels, and 20–60 axial slices per scan. Pixel spacing ranged from 0.5–1.5 mm in the X and Y directions, and 2–5 mm along the Z-axis. Additional parameters such as Echo Time (TE), Inversion Time (TI), and Repetition Time (TR) varied between 120–500 ms, 1800–2200 ms, and 2000–5000 ms, respectively. Initially, the images were provided in DICOM format, with each subject contributing around 100–200 slices. To facilitate advanced volumetric analysis, all images were converted to the Neuroimaging Informatics Technology Initiative (NifTI) format (.nii), maintaining a 1 × 1 × 1 mm voxel resolution and utilizing a 16-bit unsigned integer data structure for precision [80]. Representative MRI slices in axial, sagittal, and coronal views are shown in Figure 3.2, offering a visual understanding of the imaging data used in this study.



**Figure 3.2** NifTI image from PPMI database

### 3.3 Data Pre-Processing:

Pre-processing of MRI data is a vital preparatory step that ensures the reliability, consistency, and interpretability of radiological features used in downstream analyses. This stage transforms raw MRI scans into high-quality, normalized, and anatomically segmented images, suitable for automated machine learning pipelines and radiomics feature extraction. The pre-processing workflow developed in this study, as illustrated in Figure 3.3, consists of multiple stages including brain extraction, spatial registration, bias field correction, segmentation of subcortical regions, label fusion, and inverse registration. Each step contributes uniquely to reducing imaging artifacts, standardizing anatomical correspondence, and enhancing tissue-level resolution.



**Figure 3.3** Detailed Analysis of the pre-processing of MRI scans

### 3.3.1 Brain extraction

The initial step of MRI data pre-processing involves isolating the brain tissue from the surrounding non-brain structures within the MRI scans. For this, the Robust Brain Extraction (ROBEX) algorithm is employed. ROBEX automatically produces binary masks that separate brain tissue (labeled as 1) from non-brain areas (labeled as 0). It refines these boundaries

through repeated steps that blend morphological processing with probabilistic modelling to achieve accurate brain extraction [81]. The accuracy of this step significantly influences subsequent procedures such as registration and segmentation, making it a crucial component of the MRI analysis workflow.

### 3.3.2 Registration

To ensure anatomical consistency across all subjects, MRI scans were spatially normalized to the Montreal Neurological Institute (MNI152) standard space using the Advanced Normalization Tools (ANTs) framework. The registration pipeline included rigid transformation, followed by affine alignment and finally non-linear deformation using symmetric diffeomorphic normalization (SyN). These steps ensured precise spatial correspondence between individual subject scans and the MNI atlas [82]. Initially, affine registration corrects global geometric differences such as translation, rotation, and scaling. Following this, non-linear registration is applied to account for individual anatomical differences. The aligned images then undergo intensity normalization, which is carried out using the steps described in Eq. (3.1) to (3.3).

1. Determine the minimum and maximum intensity values within the image

$$\text{Min}_{\text{value}} = \min(I(x, y)) ; \text{Max}_{\text{value}} = \max(I(x, y)), \quad (3.1)$$

2. Subtract the minimum value from each pixel to shift the intensity range

$$\text{NormalizedImage}(x, y) = I(x, y) - \text{Min}_{\text{value}}, \quad (3.2)$$

3. Scale the image by dividing each pixel by the intensity range

$$\text{NormalizedImage}(x, y) = \frac{\text{NormalizedImage}(x, y)}{\text{Max}_{\text{value}} - \text{Min}_{\text{value}}}, \quad (3.3)$$

where  $I$  represent the T2-weighted MRI scans,  $I(x, y)$  denotes the pixel intensity at coordinates  $(x, y)$ .

### **3.3.3 Bias field correction**

Intensity inhomogeneity is a common artifact in MRI scans, typically caused by non-uniform magnetic fields and coil-related imperfections during image acquisition. These variations can obscure true tissue contrasts, adversely affecting both visual interpretation and automated analysis. To address this problem, the N4Bias Field Correction method is applied to the spatially aligned MRI images. As an enhanced successor to the N3 algorithm, N4Bias offers improved convergence, greater stability, and more effective correction of low-frequency intensity variations across the scans. This nonparametric, nonuniform intensity normalization algorithm models the bias field iteratively and applies correction, thereby improving image uniformity and preserving anatomical structures [83], [84]. Compared to N3, N4 offers increased accuracy and computational efficiency, particularly for high-resolution structural MRI, making it the preferred choice in contemporary neuroimaging pipelines. Effective bias field correction enhances segmentation precision and ensures the consistency of radiomic feature extraction across subjects.

### **3.3.4 Subcortical segmentation**

Following bias field correction, accurate segmentation of specific subcortical brain regions is performed using tools from the FMRIB Software Library (FSL). The segmentation procedure leverages probabilistic atlases and advanced ML techniques to delineate eight key subcortical regions defined by the atlas. The precision and reliability of segmentation are critical, as these brain regions significantly impact the characterization and progression assessment of PD [85].

### **3.3.5 Label fusion and Inverse registration**

Label fusion is applied to enhance segmentation accuracy. This technique combines multiple segmentations by implementing a weighted voting scheme. Specifically, each voxel's probability of belonging to a specific label is calculated, and a weighted sum is applied to generate a more reliable final segmentation, as expressed in Eq. (3.4).

$$L(v) = \arg \max_{l \in L_{atlas}} \left( \sum_{s=1}^n W_s P_s(v, l) \right) \quad (3.4)$$

where  $P_s(v, l)$  represents the likelihood that voxel  $v$  belongs to label  $l$  in segmentation output  $s$ ,  $W_s$  refers to the assigned weight for each segmentation result  $s$ , and  $L(v)$  denotes the final label attributed to voxel  $v$ . This process of combining multiple segmentations into a final consensus is visually demonstrated in Figure 3.4. Following segmentation, inverse registration is used to project the labeled regions back into the subject's original anatomical space. This backward transformation relies on the parameters initially computed during the atlas-based registration step, ensuring that the segmentation aligns precisely with each patient's unique brain structure. This method facilitates highly accurate, individualized segmentations, making the output more reliable for downstream analysis [86]. A complete overview of the MRI preprocessing workflow including each computational step is presented below as a structured pseudocode in Table 3.1.

**Table 3.1** Pseudocode of MRI pre-processing pipeline

---

### Step 1: Initialization

*Input:* NIfTi format MRI scans for all subjects

### Step 2: Brain Extraction

For each NIfTI file:

    Use ROBEX to isolate brain regions

    Generate binary masks (brain = 1, non-brain = 0)

### Step 3: Atlas Registration

For each brain-extracted image:

    Load the MNIPD25-T1MPRAGE-1 atlas

    Align the image to the atlas using ANTsPyX

### Step 4: Intensity Normalization

For each registered image:

    Calculate the min and max intensity value using Eq. (3.1)

    Normalize the pixel image intensity using Eq. (3.2 and 3.3)

### Step 5: Bias Field Correction

For each normalized scan:

    Apply the N4Bias correction algorithm

### **Step 6: Subcortical Segmentation**

For each bias-corrected image:

    Segment eight subcortical structures using FSL tools and the registered atlas

### **Step 7: Probabilistic Label Fusion**

For each group of initial segmentations:

    For each voxel:

        Estimate the probability for each label using Eq. (3.4)

        Compute final label based on weighted voting

### **Step 8: Inverse Registration to native space**

For each segmented image:

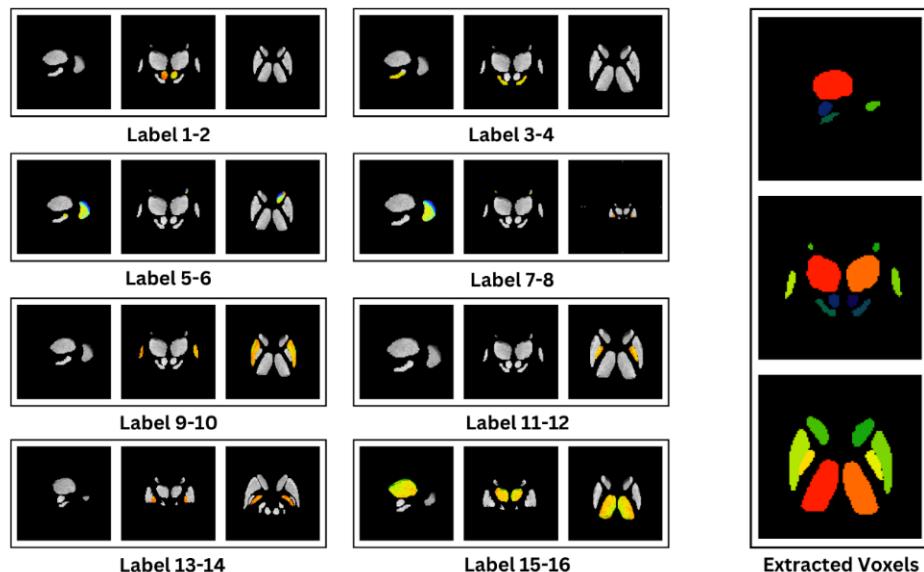
    Apply inverse registration using parameters from Step 3

    Map the labels back to the subject's original anatomical structure

End of Algorithm

*Output:* Preprocessed images with corrected bias, registered to atlas, and labeled for anatomical regions

---



**Figure 3.4** Label Fusion and Extracted Voxels from an MRI after Segmentation

### 3.4 Selection of subcortical regions and atlas-based segmentation

Accurate identification and segmentation of subcortical brain regions are crucial for detecting the subtle neuropathological alterations associated with PD. This study specifically targets deep brain structures that play pivotal roles in motor control, dopaminergic signalling, and cognitive function domains known to be significantly impaired in PD. Segmentation was performed using the MNIPD25-Subcortical-1mm atlas, a high-resolution template manually annotated by expert clinical neuroanatomists. This atlas delineates sixteen regions of interest (ROIs), encompassing bilateral volumes of eight key subcortical structures. Its fine-grained anatomical precision ensures consistent and reproducible segmentation across subjects, providing a robust foundation for radiomics-based feature extraction.

Each selected structure holds both anatomical importance and pathophysiological relevance to PD progression. The Substantia Nigra (SN), for instance, is the primary locus of dopaminergic neuron degeneration, making it a critical region of interest. The Putamen and Caudate Nucleus—components of the striatum are integral to motor regulation and demonstrate substantial neurodegenerative changes in PD. Similarly, the Globus Pallidus (both internal and external segments, GPi and GPe), Thalamus, Red Nucleus (RN), and Subthalamic Nucleus (STN) are essential elements of the basal ganglia circuitry and motor control pathways. These regions often exhibit pathological features such as volume reduction, iron accumulation, and disrupted connectivity in PD patients [87], [88]. Table 3.2 provides a summary of the anatomical labels, lateralization, and specific functional roles of each subcortical structure analyzed in this study.

**Table 3.2** Labels, Structures, and Uses of each subcortical structure in the Atlas

<b>Label</b>	<b>Subcortical Structure</b>	<b>Uses</b>
1	Left Caudate Nucleus	Involved in cognitive processing and motor planning; reduced volume has been observed in early PD.
2	Right Caudate Nucleus	Similar to the left, plays a role in learning and inhibitory control; often affected in cognitive decline.
3	Left Putamen	Key structure in motor control; undergoes dopaminergic denervation in PD.
4	Right Putamen	Mirrors the left in function; associated with rigidity and bradykinesia symptoms in PD.
5	Left Globus Pallidus Internus (GPi)	Participates in movement inhibition; altered firing patterns in PD lead to motor impairment.
6	Right Globus Pallidus Internus (GPi)	Acts as a relay for basal ganglia output; targeted in DBS for PD treatment.
7	Left Globus Pallidus Externus (GPe)	Regulates thalamocortical loops; imbalance here can exacerbate tremors in PD.
8	Right Globus Pallidus Externus (GPe)	Similar to the left, influences the indirect motor pathway; disrupted in PD
9	Left Thalamus	Relays motor and sensory signals; thalamic atrophy and hypoconnectivity are common in PD.
10	Right Thalamus	Functions as a cortical relay; associated with tremor circuitry and postural instability.
11	Left Subthalamic Nucleus (STN)	Modulates output of basal ganglia; hyperactivity linked to PD tremors and dyskinesia.
12	Right Subthalamic Nucleus (STN)	Targeted in DBS for motor symptom control; essential for rhythm and movement regulation.
13	Left Substantia Nigra (SN)	Principal site of dopamine loss in PD; shows early-stage neuromelanin reduction and iron accumulation.
14	Right Substantia Nigra (SN)	Mirrors the left SN in degeneration patterns; critical for early PD detection via imaging.
15	Left Red Nucleus (RN)	Coordinates motor activity; degeneration here may contribute to balance and gait issues in PD.
16	Right Red Nucleus (RN)	Supports limb movement coordination; structural alterations can indicate advanced PD stages.

### 3.5 Radiomics-Based Classification of PD Severity

Once subcortical structures were segmented, radiomic analysis was employed to classify the severity of PD by quantitatively characterizing the shape, intensity, and texture of these regions. Radiomics involves the automated extraction of numerous quantitative features from medical imaging data, allowing for the identification of subtle patterns that may not be easily recognized by the human eye. This method offers a clear advantage over traditional radiological assessment, which largely relies on visual interpretation and may be subject to observer variability, limited resolution, and diagnostic subjectivity. In contrast, radiomic features provide reproducible, objective, and mathematically defined descriptors of tissue heterogeneity, morphology, and internal structure essential for capturing early-stage pathological changes in PD. To ensure a standardized and reproducible feature extraction process, this chapter adopted the Pyradiomics Python library. PyRadiomics is a popular open-source tool designed for extracting radiomic features in line with the Image Biomarker Standardisation Initiative (IBSI) guidelines. It supports reproducible analysis and provides a wide range of features, including first-order statistics, shape-based metrics, and advanced texture features like GLCM and GLRLM. In this study, PyRadiomics was selected for its strong documentation, reliability in neuroimaging research, and compatibility with NIfTI-format brain MRI scans.

Using Pyradiomics, a total of 107 features were initially extracted from each of the 16 segmented subcortical regions [89]. These features spanned seven categories, including first-order intensity features, shape-based descriptors (2D and 3D), and several textural matrices. A breakdown of feature types, counts, and clinical relevance is provided in Table 3.3.

To minimize redundancy and improve model performance, a two-step feature selection strategy was implemented. First, Pearson correlation analysis was used to eliminate highly collinear features, reducing the dimensionality from 107 to 39. Then, Recursive Feature Elimination (RFE) was applied to identify the 20 most informative features, ranked using importance scores from a Random Forest classifier. RFE was specifically chosen over LASSO or Principal

Component Analysis (PCA) because it preserves the interpretability of the original features—an essential requirement for downstream explainability using SHAP and LIME. While LASSO performs embedded selection, it may discard features arbitrarily when predictors are correlated, and PCA transforms the original space into abstract components, which hinders clinical insight. In contrast, RFE provides both ranking and retention of actual features, making it ideal for transparent, model-agnostic interpretation. To address the inherent class imbalance in the dataset—particularly the underrepresentation of prodromal PD cases—the Synthetic Minority Over-sampling Technique (SMOTE) was applied. SMOTE synthetically generates new instances for minority classes by interpolating between existing samples, thereby producing a more balanced and statistically robust training set [90]. A schematic overview of the grading and feature selection process is shown in Figure 3.5.

**Table 3.3** Explanation of features type and number of features extracted from ROI

Feature Type	Number of Features	Explanation
First-order Statistical Features	17	Capture basic intensity statistics such as mean, variance, skewness indicative of tissue density changes in PD.
Shape-Based (2D)	10	Describe two-dimensional geometric properties like area and perimeter; useful for detecting regional atrophy.
Shape-based (3D)	10	Capture volumetric and surface characteristics, crucial for assessing structural shrinkage.
Gray Level Cooccurrence Matrix	24	Analyze spatial relationships between pixel intensities, useful for detecting microstructural disruptions.
Gray Level Run Length Matrix	12	Evaluate length patterns of identical intensities; sensitive to texture regularity.
Gray Level Size Zone Matrix	12	Assess size of homogenous intensity zones; helpful for identifying uniform tissue loss.
Neighbouring Gray Tone Difference Matrix	12	Measure differences in local intensity gradients; reflects textural complexity.

Table 3.3 (Continues)

Gray Level Dependence Matrix	10	Quantify dependence between voxels; useful in detecting structural symmetry or irregularity.
Exhaustive set of Features (Total)	107	Comprehensive set enabling deep characterization of tissue properties in PD.

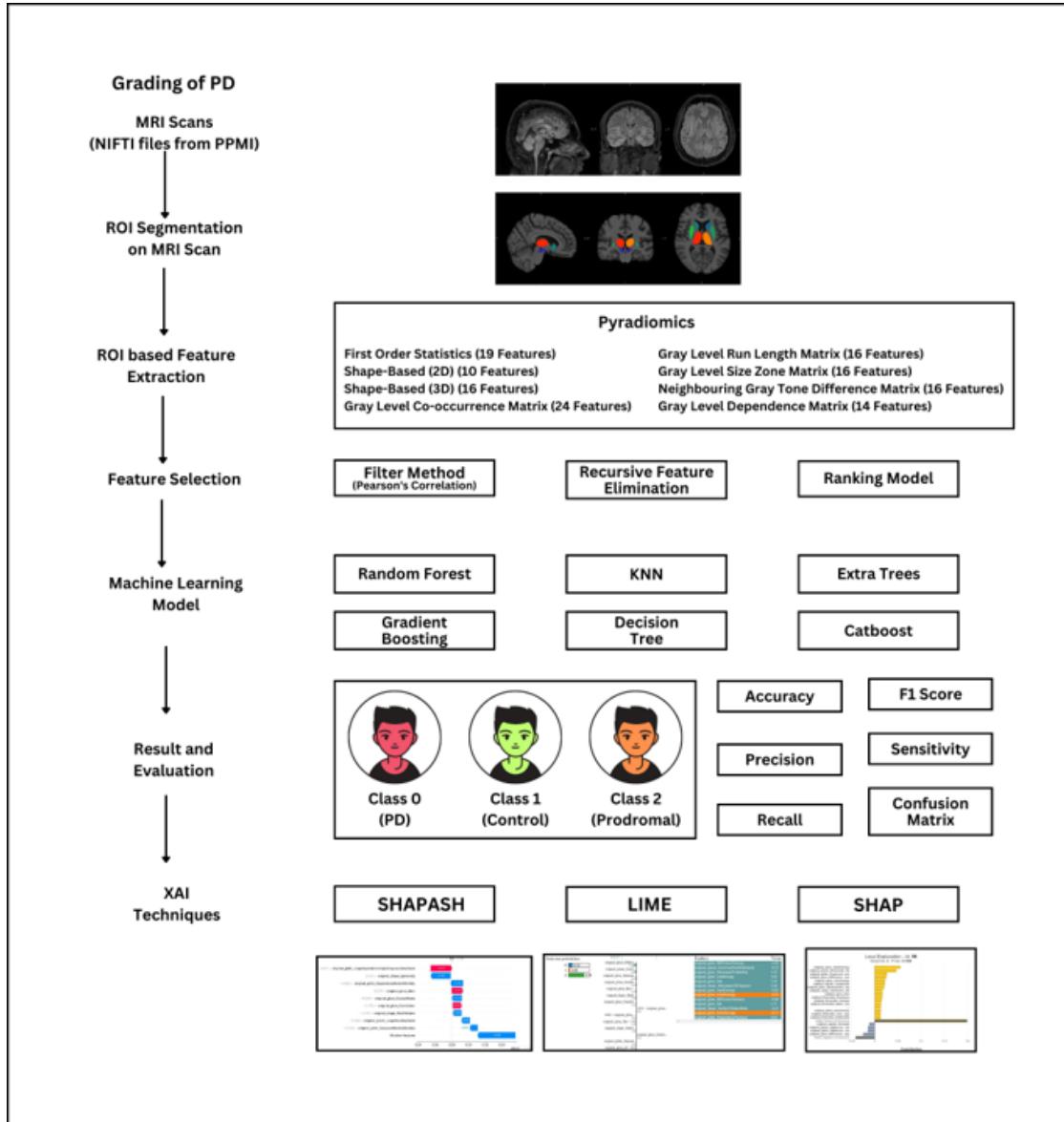
### 3.6 ML Classifier

ML classifiers were developed to accurately categorize subjects into three distinct groups: PD, HC, and prodromal cases. This classification aimed to facilitate the early detection of PD through automated processes based on radiomic features derived from MRI scans. Six popular classifiers were evaluated: RF, K-Nearest Neighbors (KNN), Extra Trees (ET), Gradient Boosting (GB), Decision Tree (DT), and CatBoost. The top 20 radiomic features selected during the feature selection phase were utilised as inputs for these models. Classifier performance was rigorously assessed using various standard evaluation metrics, including accuracy, precision, recall, F1-score, specificity, sensitivity, and confusion matrices. Cross-validation techniques were employed further to validate the robustness and generalizability of the models. Receiver Operating Characteristic (ROC) curves, along with Area Under the Curve (AUC) scores, were computed to assess each classifier's discriminative capability for different classes. Additionally, calibration curves were generated to evaluate the reliability of the predicted probabilities, an essential aspect of clinical decision-making.

### 3.7 XAI

Incorporating XAI into the ML framework is fundamental for enhancing transparency and interpretability in clinical decision-making processes. In this research, several XAI methods were explored, including SHAP, LIME, and SHAPASH, each providing different insights into model decisions. SHAP assigns a global importance value to each feature, clearly demonstrating its overall contribution to the predictive model. SHAP values are computed based on Eq. (3.5):

$$SHAP(x) = \emptyset_0 + \sum_{i=1}^M \phi_i(x) \quad (3.5)$$



**Figure 3.5** Overview of the complete workflow designed for PD grading.

where  $\emptyset_0$  represents the base value or initial prediction, and  $\phi_i(x)$  denotes the SHAP value corresponding to feature  $i$  for a given instance  $x$ . SHAP visualizations, including waterfall and beeswarm plots, are particularly useful for explaining model predictions—highlighting both the

contribution of individual features to a specific outcome and their overall significance across the dataset [91].

LIME provides local explanations by analyzing the variations in predictions when input data is altered slightly, thus enabling clinicians to understand model predictions on a case-by-case basis [92]. The mathematical formulation of LIME is given by Eq. (3.6):

$$\xi(x) = \arg \min_{g \in G} L(f, g, w_x) + w(g) \quad (3.6)$$

where  $\xi(x)$  represents the interpretable surrogate model for a specific instance  $x$ , while  $G$  denotes the set of possible explanations. The term  $L$  refers to the loss function that quantifies how well the explanation model approximates the original model  $f$ . The weighting factor  $w_x$  is applied to each sampled instance, emphasizing its influence based on how representative or important it is within the original dataset. When the sampled instance closely aligns with the original data distribution,  $w_x$  receives a higher value. SHAPASH builds upon this by offering enhanced interpretability through rich visualizations and summary reports, helping to clearly communicate how different features contribute to model predictions and their variability using SHAP values. These XAI methodologies collectively enhance model interpretability, enabling clinicians to better understand and trust machine-generated predictions, ultimately facilitating personalized and effective patient care.

### **3.8 Result and analysis of ML classifiers**

This study carried out a comparative evaluation using multiple machine learning classifiers to analyze textural, morphological, and statistical features extracted from segmented subcortical MRI regions. The detailed configurations and parameter settings for each classifier are presented in Table 3.4. For consistency and reproducibility, CatBoost (v1.2.5) and Scikit-learn (v1.2.2) were used throughout the experiments. The radiomic features were computed from a dataset of 500 segmented MRI scans, which was split into 80% for training and 20% for testing. Initially, classifier performance was hindered by class imbalance. To correct this, the Synthetic

Minority Oversampling Technique (SMOTE) was applied, resulting in improved class distribution and enhanced model performance, as shown in Table 3.5. Figure 3.6 illustrates the feature space before and after SMOTE application, clearly depicting better class separation and balance.

**Table 3.4** Hyperparameter Settings for Various ML Classifiers Used in the Study

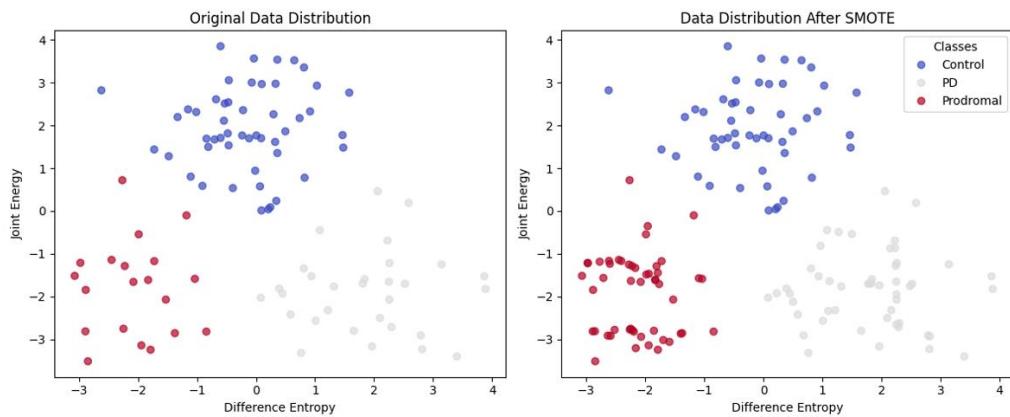
ML classifier models	Specifications
Random Forest classifier	n_estimators: [500, 1000], max_depth: [None, 10, 20], criterion: ['gini', 'entropy'], min_samples_split: [2, 5], max_features: ['sqrt', 'log']
KNN classifier	n_neighbors: [3, 5, 7], weights: ['uniform', 'distance'], p: [1, 2], algorithm: 'auto', leaf_size: [20, 30]
Extra Trees Classifiers	n_estimators: [500, 1000], max_depth: [None, 10, 20], criterion: 'gini', min_samples_split: [2, 5], max_features: ['sqrt', 'log']
Gradient Boosting Classifier	n_estimators: [500, 1000], learning_rate: [0.1, 0.5], max_depth: [3, 5], loss: 'log_loss', min_samples_split: [2, 4], min_samples_leaf: [1, 5]
Decision Tree Classifier	max_depth: [None, 10, 20], criterion: 'gini', splitter: 'best', min_samples_split: 2, min_samples_leaf: 1
Catboost Classifier	iterations: [100, 200], learning_rate: [0.1, 0.5], depth: [3, 5], loss_function: 'log_loss'

**Table 3.5** Comparison of classifier accuracy before and after applying SMOTE

Model	Accuracy before using SMOTE	Accuracy after using SMOTE
Random forest	76.6%	89.2%
KNN	74.3%	87.8%
Extra Trees	84.7%	93.6%
Gradient Boosting	86%	96.8%
Decision Trees	81.2%	90%
Catboost	78.7%	92%

### 3.9 Performance Evaluation and Comparative Analysis

A thorough evaluation of various ML classifiers, including RF, KNN, ET, GB, DT, and CatBoost, was performed using five distinct test splits. Evaluation metrics included accuracy, precision, recall, specificity, F1 score, and computation time, as detailed in Table 3.6.



**Figure 3.6** Scatter plot visualising feature distribution before and after applying SMOTE, highlighting class separation improvement.

These metrics, computed using Eq. (3.7) to (3.11), provide critical insights into the performance and reliability of the classifiers.

$$Accuracy = \frac{(True\ Positive + True\ Negative)}{(TP + FP + True\ Negatives + False\ Negatives)} \quad (3.7)$$

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)} \quad (3.8)$$

**Table 3.6** Performance comparison of various ML classifiers used in PD grading

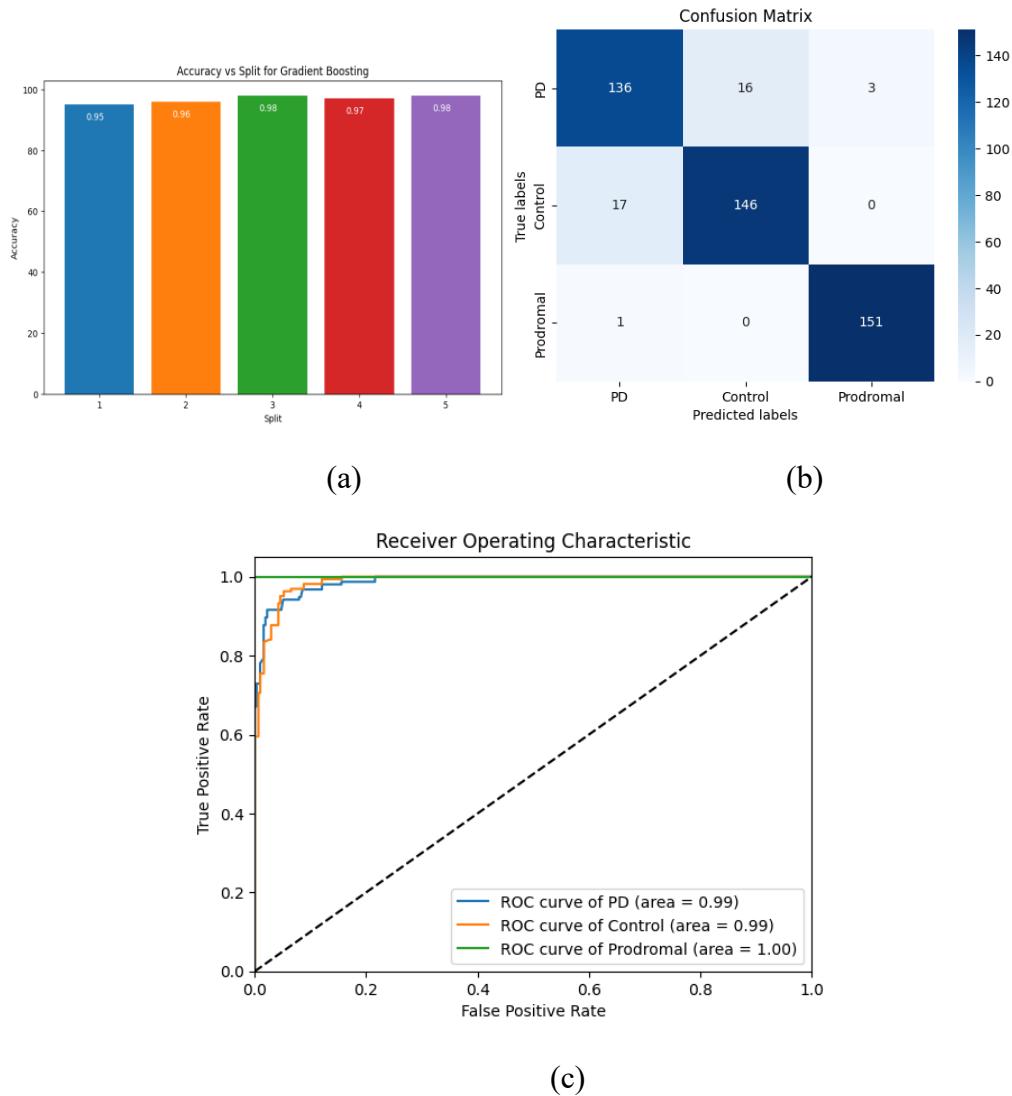
Model	Split	Accuracy (in %)	Precision (in %)	Recall (in %)	Specificity (in %)	F1-score (in %)	elapsed-time (in sec)
Random forest	Split 1	89	90	89	88	91	90.23
KNN		88	89	79	89	84	53.68
Extra Trees		93	96	87	95	92	56.92
Gradient		95	96	92	95	94	78.63
Boosting		87	85	76	87	80	41.72
Decision Trees		90	94	84	93	88	43.86
Catboost							
Random forest	Split 2	88	90	81	81	91	82.54
KNN		89	94	83	93	85	34.91
Extra Trees		92	94	84	93	92	42.25
Gradient		96	98	92	97	95	66.04
Boosting		89	90	89	88	90	30.84
Decision Trees		92	90	86	92	88	26.46
Catboost							
Random forest	Split 3	90	96	82	95	98	80.25
KNN		87	85	76	89	79	41.45
Extra Trees		94	96	89	95	92	42.91
Gradient		98	98	98	97	95	83.51
Boosting		90	92	83	90	87	31.05
Decision Trees		91	92	83	91	87	33.29
Catboost							
Random forest	Split 4	89	94	83	93	88	92.33
KNN		88	89	79	89	86	40.69
Extra Trees		96	98	92	97	94	49.84
Gradient		97	95	91	97	91	72.04
Boosting		92	84	93	88	88	30.68
Decision Trees		93	96	85	95	90	33.57
Catboost							
Random forest	Split 5	90	95	81	95	92	79.42
KNN		87	84	76	88	80	40.67
Extra Trees		93	94	89	93	91	59.39
Gradient		98	98	98	97	98	83.72
Boosting		92	81	90	80	85	30.73
Decision Trees		94	98	87	97	92	33.45
Catboost							

$$Recall = \frac{\text{True Positives}}{(\text{True Positives} + \text{False Negatives})} \quad (3.9)$$

$$Specificity = \frac{\text{True Negatives}}{(\text{False Positives} + \text{True Negatives})} \quad (3.10)$$

$$F1\ score = \frac{(2 * \text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (3.11)$$

GB emerged as the most effective model, achieving the highest accuracy (96.8%), precision (97%), recall (94.2%), specificity (96.6%), and F1-score (94.6%) for PD detection. Although DT had the fastest computational time (approximately 33.0 seconds) due to simpler node-splitting mechanics, GB provided an optimal balance of accuracy and computational efficiency, averaging a processing time of 76.78 seconds. RF, although effective, required longer computation times, averaging 84.9 seconds, due to its ensemble structure. Figure 3.7 (a) depicts the performance stability of the GB classifier across five-fold cross-validation splits, consistently outperforming other classifiers. The confusion matrix illustrated in Figure 3.7 (b) underscores the model's effectiveness, particularly in identifying prodromal PD cases accurately, which is essential for early-stage detection. Furthermore, the ROC curves, shown in Figure 3.7 (c), demonstrate exceptional discriminative performance with AUC values of 0.99 for PD and control groups, and a perfect AUC of 1.00 for the prodromal group. Despite its predictive robustness, the inherent 'black box' nature of the GB model limits clinical transparency and interpretability. To overcome this limitation and promote clinical adoption, integrating XAI methodologies is crucial. XAI provides meaningful insights into decision-making processes, ensuring ML model predictions are both effective and comprehensible, thereby enhancing clinician trust and facilitating personalized patient care.



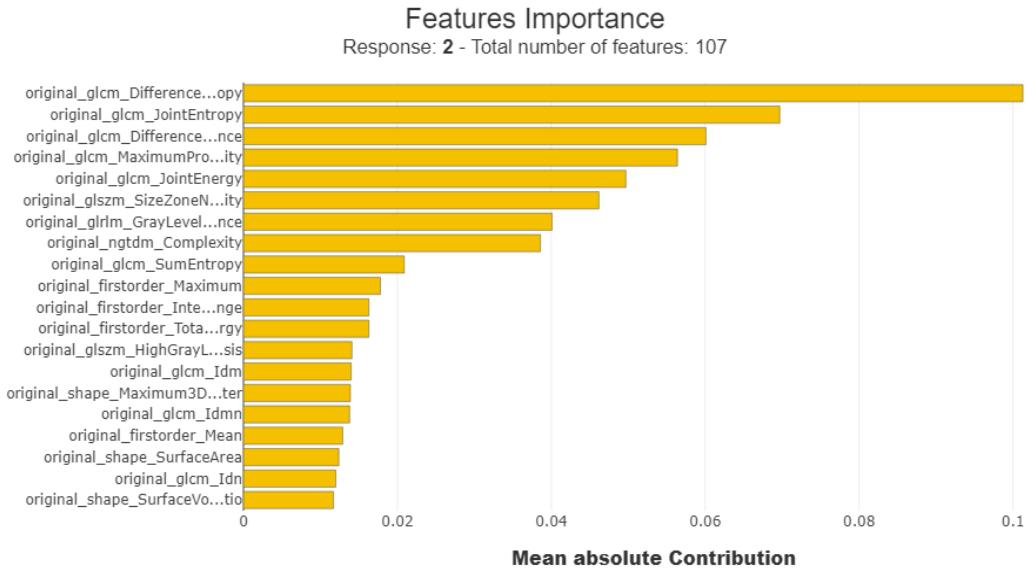
**Figure 3.7** (a) Accuracy trends of the Gradient Boosting classifier across 5-fold cross-validation splits; (b) Confusion matrix highlighting prediction performance of the Gradient Boosting model; (c) ROC curves illustrating classification performance for the three PD-related classes.

### 3.10 Analysis and Interpretation Using XAI Techniques

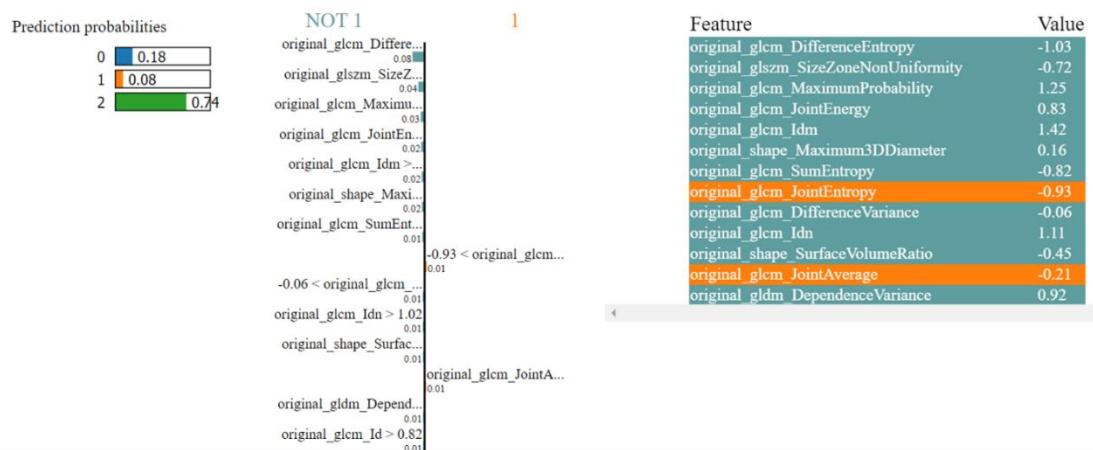
To improve the clarity and trustworthiness of AI-based diagnostic models, this study integrated Explainable AI techniques—specifically SHAP and LIME. These methods help uncover which features have the greatest impact on the model’s decisions, offering insights both at a broader level (across all subjects) and at an individual case level. As illustrated in Figure 3.8(a), the Feature Importance Plot highlights the most influential features contributing to predictions for

the prodromal group. Notably, features such as Difference Entropy, Joint Entropy, and Difference Variance emerged as key indicators. From a clinical standpoint, higher values of Difference Entropy often signal increased tissue irregularity—an early sign of neurodegeneration seen during the prodromal phase of PD. This alignment with established clinical patterns lends credibility to the model's interpretability and decision-making process. The local interpretation generated using LIME, as shown in Figure 3.8(b), reveals how the model arrived at its prediction: assigning an 18% chance to PD, 8% to the control group, and 74% to the prodromal stage. In this particular case, features like Difference Entropy and Joint Entropy strongly contributed to identifying the prodromal condition, while Surface Volume Ratio had a negative influence on that classification. This detailed level of explanation aligns well with clinical understanding and helps healthcare professionals interpret individual predictions with greater clarity and trust.

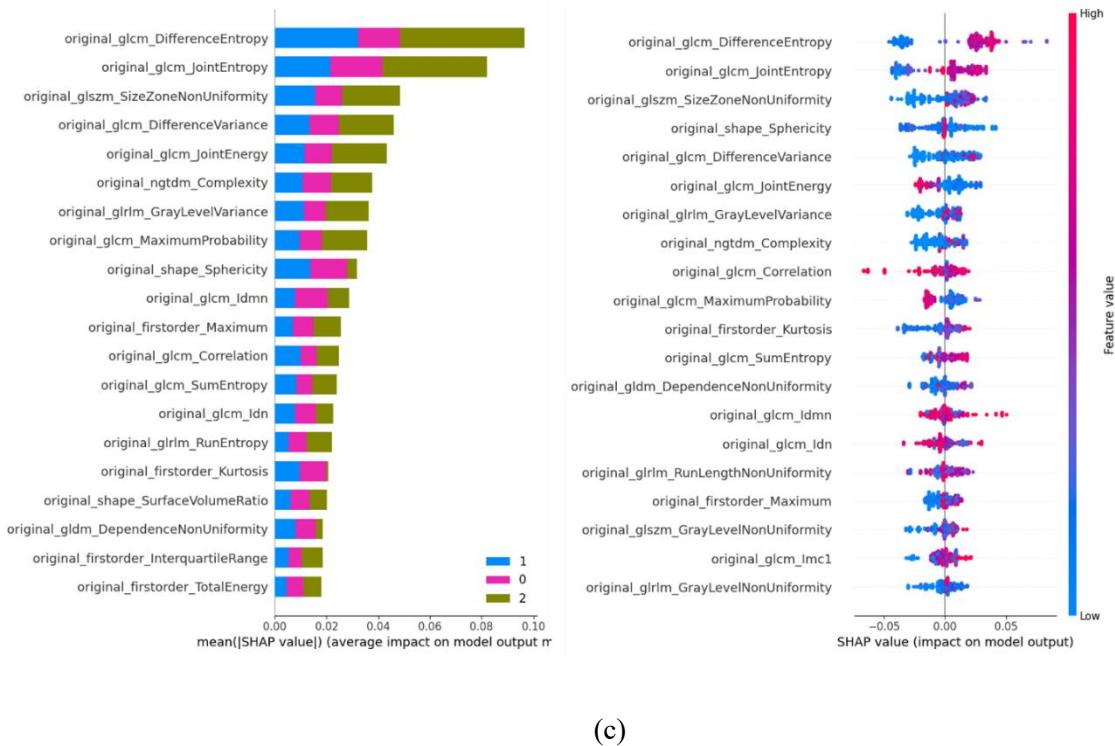
The SHAP summary plot in Figure 3.8(c) offers a global overview of feature contributions across the entire dataset. It confirms the consistent importance of Difference Entropy and Joint Entropy in distinguishing disease stages. Notably, higher values of Difference Entropy were consistently associated with prodromal and early PD classes, suggesting these features capture physiologically meaningful tissue disorganization patterns relevant to early neurodegenerative processes. While XAI tools greatly enhance interpretability, certain limitations must be acknowledged. For instance, LIME's explanations are locally linear approximations and may become unreliable if the model's decision surface is highly non-linear in the vicinity of a prediction. Similarly, SHAP interpretations depend on feature independence assumptions, which may not always hold in high-dimensional clinical data. Overall, the integration of SHAP and LIME provides a balanced combination of local and global interpretability. These tools empower clinicians to better understand and trust AI-based diagnostic models, thereby facilitating more transparent, evidence-backed decision-making in real-world healthcare settings.



(a)



(b)



**Figure 3.8** Interpretability analysis of the GB classifier for PD classification (a) SHAPASH highlights feature importance at the individual instance level; (b) LIME provides localized explanations for a specific prediction; (c) SHAP visualizes both global and local contributions of each feature, offering detailed insights into the model's decision process.

### 3.11 Statistical Analysis

In this study, thorough statistical testing was carried out to validate the radiomic features extracted from MRI scans for their usefulness in identifying early signs of PD, including its prodromal phase. As shown in Table 3.7, twenty key radiomic features were examined to assess their ability to distinguish between healthy controls, PD patients, and those at risk. The analysis began with ANOVA, which revealed statistically significant differences ( $p < 0.05$ ) across all groups indicating the strong diagnostic potential of these features. Given the complexity and variability of medical imaging data, which may not always meet the assumptions of normality or equal variance, a non-parametric Kruskal-Wallis H-test was also performed. This test supported the ANOVA results, confirming that the selected features remain effective in distinguishing between disease stages even when data distributions are not ideal. Together,

these statistical validations highlight the robustness and clinical relevance of the features used for early Parkinson's detection.

**Table 3.7** Statistical analysis of the selected features

Selected features	PD	Prodromal	Control	p-value	Significant
Maximum 2D Diameter Column (shape)	0.33 ± 0.07	0.54 ± 0.56	1.09 ± 1.63	1.17	True
Maximum2DDiameterRow (shape)	0.31 ± 0.13	0.84 ± 0.40	1.12 ± 1.56	6.07	True
Maximum2DDiameterSlice (shape)	0.33 ± 0.09	0.82 ± 0.09	1.03 ± 1.60	2.75	True
Maximum3DDiameter (shape)	0.32 ± 0.07	0.53 ± 0.06	1.06 ± 1.57	3.55	True
Surface Area (shape)	0.07± 0.72	0.85 ± 0.76	0.89 ± 1.12	6.16	True
90Percentile (first-order)	0.15± 0.18	0.45 ± 0.19	1.13 ± 0.15	9.58	True
Median (first-order)	0.13 ± 1.85	0.30 ± 0.16	1.18 ± 0.16	9.19	True
Range (first-order)	0.14 ± 1.84	0.30 ± 0.28	1.19 ± 0.19	3.73	True
Difference Entropy (glcm)	0.19 ± 0.86	0.23 ± 0.78	1.04 ± 0.77	3.89	True
Idn (glcm)	-0.21±1.09	-0.41±1.10	0.68 ± 1.03	8.64	True
Joint Entropy (glcm)	0.13± 0.63	0.55 ± 0.71	1.04 ± 1.06	1.88	True
Maximum Probability (glcm)	-0.26±0.71	-0.51±0.54	1.24 ± 1.10	1.22	True
Dependence Entropy (gldm)	0.14± 0.79	0.37 ± 0.68	-0.68±1.28	6.63	True
Gray Level Non-Uniformity (gldm)	-0.06±0.74	-0.21±0.36	0.37 ± 1.18	1.08	True
Gray Level Non-Uniformity (glrlm)	-0.07±0.71	-0.19±0.37	0.33 ± 1.12	2.82	True
Gray Level Non-Uniformity Normalized(glrlm)	-0.11±0.92	-0.47±0.66	0.87 ± 1.05	2.82	True
Run Variance (glrlm)	0.06 ± 1.12	-0.22±0.73	0.40 ± 1.29	3.82	True
Gray Level Non-Uniformity Normalized (glszm)	-0.19±0.89	-0.43±0.66	0.75 ± 0.94	4.04	True
Zone Entropy (glszm)	-0.12±0.87	-0.3 ± 0.68	0.62 ± 1.31	1.36	True
Busyness (ngtdm)	-0.09±0.73	-0.23±0.53	0.45 ± 1.14	2.25	True

### 3.12 Discussion on Radiomic and XAI Insights

Radiomic features used in this research provide significant biological insights into PD progression. Shape-based features, such as Maximum Diameter and Surface Area, are effective indicators of subcortical structural shrinkage and neuronal loss, typical in PD pathology. Features assessing image intensity variations, influenced by substances such as iron, offer additional diagnostic markers due to their correlation with disease severity [8], [93]. Texture-based features like Gray Level Non-Uniformity and its Variance are indicative of tissue uniformity changes, providing insights into PD-induced alterations in brain tissue architecture [94], [95]. The physical interpretation of the selected imaging features provides significant insights into the progression of PD:

- (a) Shape-based features (e.g., Maximum 2D and 3D diameters, Surface Area): Reduced values in PD cases imply substantial structural shrinkage linked to neuron degeneration. Intermediate values in prodromal subjects suggest early structural alterations, while larger measurements in control groups reflect normal morphology.
- (b) First-order statistical features (Median, 90th Percentile, Range): Lower values in PD patients indicate reduced intensity due to neuronal loss in affected regions, with prodromal cases showing intermediate values, suggesting early-stage changes.
- (c) Texture analysis (GLSZM and NGTDM): Texture-related features such as entropy and maximum probability reflect degeneration and tissue uniformity. Lower entropy and higher maximum probability values in PD patients indicate less complex, more uniform textures, symptomatic of degenerative processes. Intermediate values in prodromal cases reflect early-stage textural changes, while control subjects display higher complexity and healthier tissue structures.
- (d) Texture parameters relating to BG size (e.g., Entropy, Maximum Probability, GLSZM): Higher entropy denotes greater complexity and heterogeneity, whereas maximum probability

highlights negative uniformity indicative of pathological conditions. These texture measures provide crucial insights into early pathological transformations.

Integrating XAI techniques into the GB classifier enhances transparency and clinical acceptance. Both local explanations (LIME) and global explanations (SHAP and SHAPASH) demonstrated the significant role of Difference Entropy and Joint Entropy features, validating their critical contribution to predictive accuracy. Such integration not only reinforces clinical trust but also supports personalised patient care through actionable, interpretable insights.

### **3.13 Summary**

This study presents a robust Explainable XAI-driven framework for the early detection and classification of PD using MRI-based radiomic features from subcortical brain regions.

- A detailed preprocessing pipeline was designed, including brain extraction, atlas-based registration, bias field correction, and segmentation of 16 subcortical structures, ensuring high-quality input data for analysis.
- Through an extensive evaluation of multiple ML classifiers, GB emerged as the best-performing model, achieving a classification accuracy of 96.8%, supported by metrics such as precision, recall, F1-score, and ROC-AUC.
- The integration of SHAP and LIME provided both global and local interpretability, making the model's predictions more transparent and clinically trustworthy by highlighting critical features like Difference Entropy and Joint Entropy.
- Statistical validation through ANOVA and Kruskal-Wallis tests confirmed the significance of selected radiomic features in distinguishing PD, prodromal, and healthy controls, reinforcing the clinical relevance of the model.

- Overall, the proposed framework combining radiomics, ML, and XAI offers a powerful and interpretable tool for early PD detection, paving the way for more personalised and explainable diagnostic solutions in neuroimaging.

## CHAPTER 4: HYBRID DEEP LEARNING FRAMEWORK FOR PARKINSON'S DETECTION

### 4.1 Introduction

PD is recognised as the second most prevalent neurodegenerative disorder globally, impacting more than 10 million individuals, with prevalence expected to significantly increase by 2040 owing to demographic aging [96]–[98]. This progressive disease predominantly manifests through motor symptoms, including tremors, slowness of movement (bradykinesia), and muscular rigidity, eventually advancing to non-motor complications affecting cognitive and emotional functions [15]. Early and accurate detection of PD is critically important as it facilitates timely intervention, potentially delaying disease progression and enhancing the quality of life for affected individuals. Nevertheless, diagnosing PD accurately in its prodromal stages remains a considerable clinical challenge due to symptom overlap with other neurological disorders, leading to diagnostic uncertainty [16], [43]. Consequently, neuroimaging techniques have gained prominence as valuable tools for improving diagnostic reliability and understanding underlying neuropathology.

Magnetic Resonance Imaging (MRI) stands out among neuroimaging methods due to its non-invasive nature and superior spatial resolution, enabling detailed structural and microstructural assessment of brain alterations associated with PD. Traditionally, research has concentrated on targeted regions such as the substantia nigra (SN) and basal ganglia (BG) due to their central involvement in dopaminergic neuron degeneration [6], [8]. However, this region-specific approach might overlook broader and more subtle neurodegenerative patterns occurring across the entire brain, which are crucial for comprehensive PD characterization. Thus, recent scientific discourse advocates for a holistic whole-brain imaging perspective to better capture the extensive and spatially dispersed pathological changes in PD. Specifically, T2-weighted MRI sequences have demonstrated significant effectiveness in visualizing subtle pathological

changes due to their enhanced tissue contrast and distinct anatomical delineation [9]. This approach allows for comprehensive analysis, identifying meaningful biomarkers potentially missed by localized assessments. Advances in AI, especially ML, have revolutionized medical image analysis by automating feature extraction, classification, and identifying diagnostic patterns [99], [100]. Yet, conventional ML techniques predominantly rely on manually engineered features, which are susceptible to biases, limited generalizability, and difficulties in capturing intricate spatial interactions within volumetric medical imaging data.

DL, particularly CNNs, overcomes these challenges by automatically learning layered feature representations directly from raw image inputs [101]. While 2D CNN methods have demonstrated efficacy in medical imaging, their application to three-dimensional MRI data results in the loss of essential volumetric information. This limitation has spurred the development of 3D-CNN models, which maintain volumetric coherence, though these methods face challenges such as higher computational demands and data scarcity, particularly with limited labelled datasets [102]. Addressing these critical gaps, this chapter presents an innovative hybrid deep-learning framework employing both custom-designed 3D-CNN and enhanced 3D Residual Network (3D-ResNet) architectures. The hybrid approach capitalizes on complementary features extracted by these distinct models and integrates them through Canonical Correlation Analysis (CCA). Further, to optimize the high-dimensional feature set, the Whale Optimization Algorithm (WOA) is utilized, enhancing predictive performance and computational efficiency. Overall, the contributions of this chapter include: (i) developing a novel 3D-CNN architecture tailored for automated feature extraction from whole-brain MRI data, (ii) enhancing and validating a 3D-ResNet architecture, (iii) effectively fusing complementary features using CCA, and (iv) optimizing the fused feature set through WOA for robust and accurate early PD diagnosis. These advancements collectively provide significant improvements over existing methodologies, offering an effective and generalizable model for clinical applications.

## 4.2 Data collection and pre-processing

The dataset used in this study was from the PPMI database ([www.ppmi-info.org/access-data-specimens/download-data](http://www.ppmi-info.org/access-data-specimens/download-data)), renowned for its comprehensive nature, public accessibility, and adherence to standardized image acquisition protocols. A total of 303 T2-weighted MRI scans were specifically chosen, consisting of 110 healthy controls, 58 individuals with prodromal symptoms, and 135 confirmed PD patients. The selected scans ensured demographic diversity, considering variations in age and gender, crucial for enhancing the model's applicability to real-world clinical scenarios. T2-weighted MRI sequences were selected over T1 and FLAIR sequences due to their distinct advantages in identifying subtle pathological changes relevant to PD diagnosis. Specifically, T2-weighted imaging provides superior tissue contrast, which is essential for effectively delineating anatomical boundaries and visualizing neurodegenerative alterations in the brain.

The preprocessing phase involved several critical steps: brain extraction to isolate relevant brain structures, intensity normalization to standardize MRI signal intensities, and data augmentation techniques aimed at enhancing dataset robustness. These preprocessing steps followed established protocols described in prior research, [103], [104] ensuring consistency and enhancing the quality of features derived from the MRI data. To maximize the framework's adaptability to diverse clinical settings, MRI scans acquired from different systems (specifically 1.5T and 3T scanners) were included. Table 4.1 provides a detailed comparison of the performance and quality characteristics associated with different MRI machines, further emphasizing the generalizability and reliability of the proposed framework across varied clinical imaging conditions.

**Table 4.1** Comparison of MRI Systems Based on Performance and Image Quality Characteristics

<b>MRI Machine</b>	<b>Manufacturer</b>	<b>Magnetic Field Strength (T)</b>	<b>Output Format</b>	<b>DICOM Image Quality Factors</b>
MAGNETOM Skyra	Siemens Healthineers	3.0	DICOM	Delivers high spatial detail, strong signal-to-noise ratio (SNR), and effective artifact suppression
MAGNETOM Aera	Siemens Healthineers	1.5	DICOM	Offers moderate resolution with reliable SNR and decent artifact control
SIGNA Premier	GE Healthcare	3.0	DICOM	Produces detailed images with excellent SNR and supports advanced imaging protocols
Discovery MR750	GE Healthcare	3.0	DICOM	Known for sharp image clarity, high SNR, and strong artifact minimization
Ingenia Elition	Philips Healthcare	3.0	DICOM	High-quality imaging with fine resolution, enhanced SNR, and integrated noise reduction
Achieva	Philips Healthcare	1.5	DICOM	Generates moderate resolution images with good SNR and improved contrast capabilities

### 4.3 Deep learning architecture

With recent advancements in computational resources, sophisticated algorithms, and substantial data storage capabilities, AI techniques have become increasingly pivotal for the early and accurate diagnosis of PD. Despite significant progress, diagnosing PD accurately during its initial stages and reliably evaluating disease severity remain challenging due to the complexities inherent in neurological disorders. Traditional ML and 2D-CNNs, widely used in prior studies, often fall short in effectively processing volumetric MRI data due to loss of critical spatial information and inconsistency caused by varying image slices [105]–[107]. To overcome these limitations, the proposed framework systematically integrates two complementary DL architectures specifically designed to harness the advantages of volumetric data processing. As

illustrated in Figure 4.1, the methodology comprises four essential modules: (1) a custom-developed 3D-CNN architecture tailored explicitly for volumetric MRI analysis, (2) an enhanced 3D-ResNet to improve feature extraction and mitigate gradient-related challenges, (3) a feature fusion strategy employing CCA, and (4) feature refinement using the WOA.

### **4.3.1 Proposed 3D-CNN**

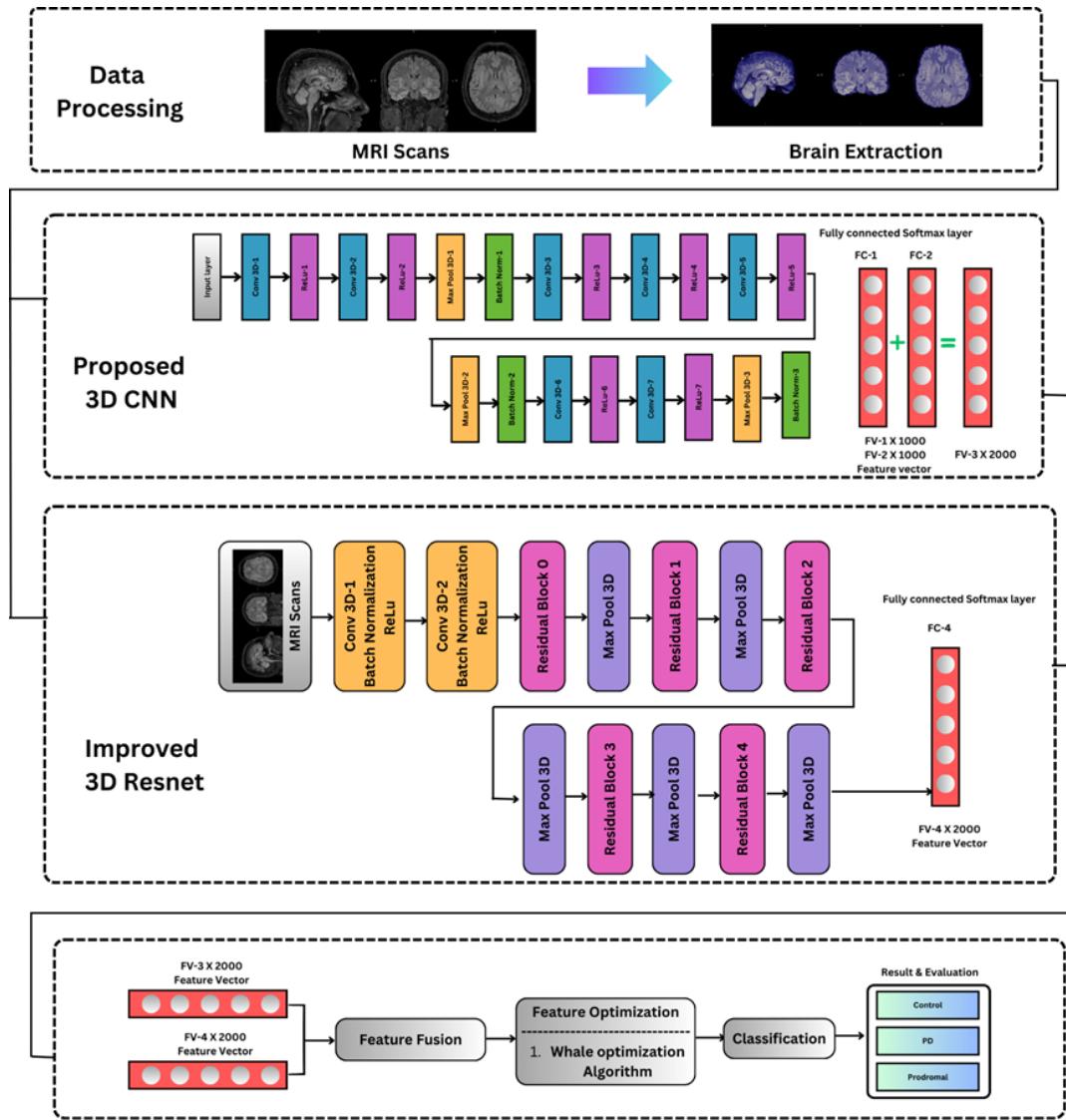
The custom 3D-CNN architecture was specifically developed to directly utilize three-dimensional MRI data in Neuroimaging Informatics Technology Initiative (NIfTI) format, effectively overcoming the limitations posed by traditional 2D-CNNs that rely on individual MRI slices, resulting in accuracy constraints due to variability in slice numbers per subjects. Extensive ablation studies guided the design of this innovative 3D-CNN, optimizing the network to accurately classify MRI images into three distinct categories: healthy controls (0), prodromal PD subjects (1), and confirmed PD patients (2). The developed 3D-CNN model consists of 24 layers, including an input layer, several 3D convolutional layers activated by ReLU functions, batch normalization layers, pooling operations, two fully connected (dense) layers, and a final SoftMax layer for classification. The complete architecture of the proposed 3D-CNN is illustrated in Figure 4.2.

### **4.3.2 Improved 3D-ResNet**

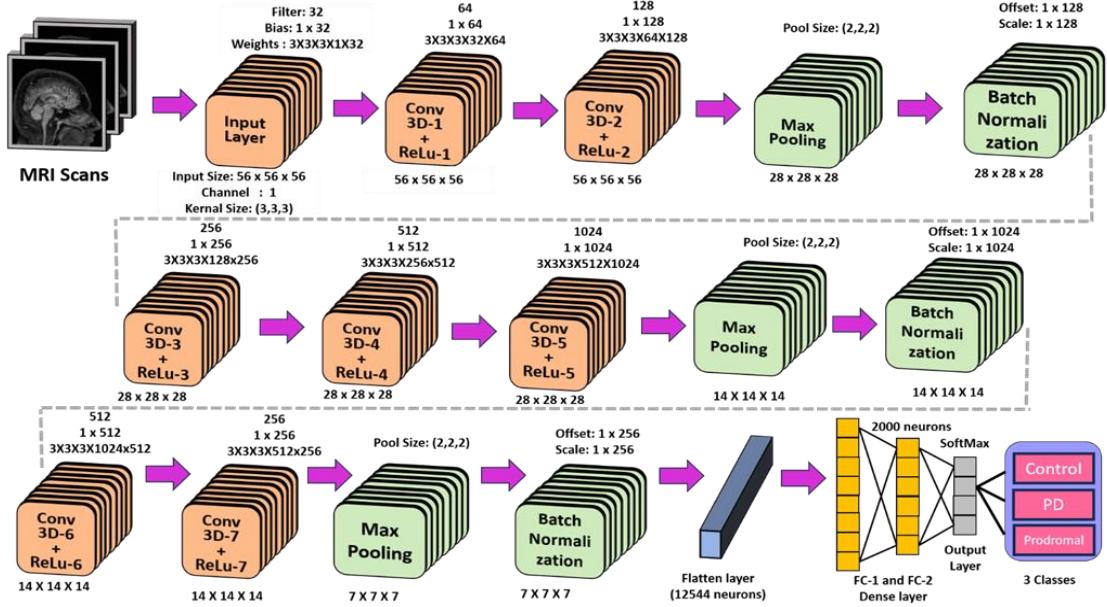
Complementing the 3D-CNN, an improved 3D-ResNet architecture was incorporated into the hybrid framework. The selection of the ResNet-based model addresses the inherent challenges of deeper network architectures, such as vanishing and exploding gradients, by employing residual connections (skip connections) that effectively propagate gradients through deep networks [55]. The improved 3D-ResNet architecture employed in this study is structured into fifteen layers, consisting of an input layer, two conv3D blocks, five residual blocks each containing max-pooling layers, followed by a fully connected layer and a SoftMax output layer. Each conv3D block comprises 3D convolutional layers, BN layers, and ReLU activations. The residual blocks are constructed from multiple residual units, each unit integrating two 3D

convolutional layers, BN layers, and ReLU activations, crucially enhanced by skip connections that enable direct gradient flow between layers, effectively alleviating gradient descent issues.

The detailed structure is illustrated in Figure 4.3.



**Figure 4.1** Overall architecture of the proposed method



**Figure 4.2** Proposed 3D-CNN

The optimization of the improved 3D-ResNet employs the Adam optimizer alongside the categorical cross-entropy loss function in its output layer, and uses ReLU activations, defined mathematically in Eq. 4.1, to enhance learning stability:

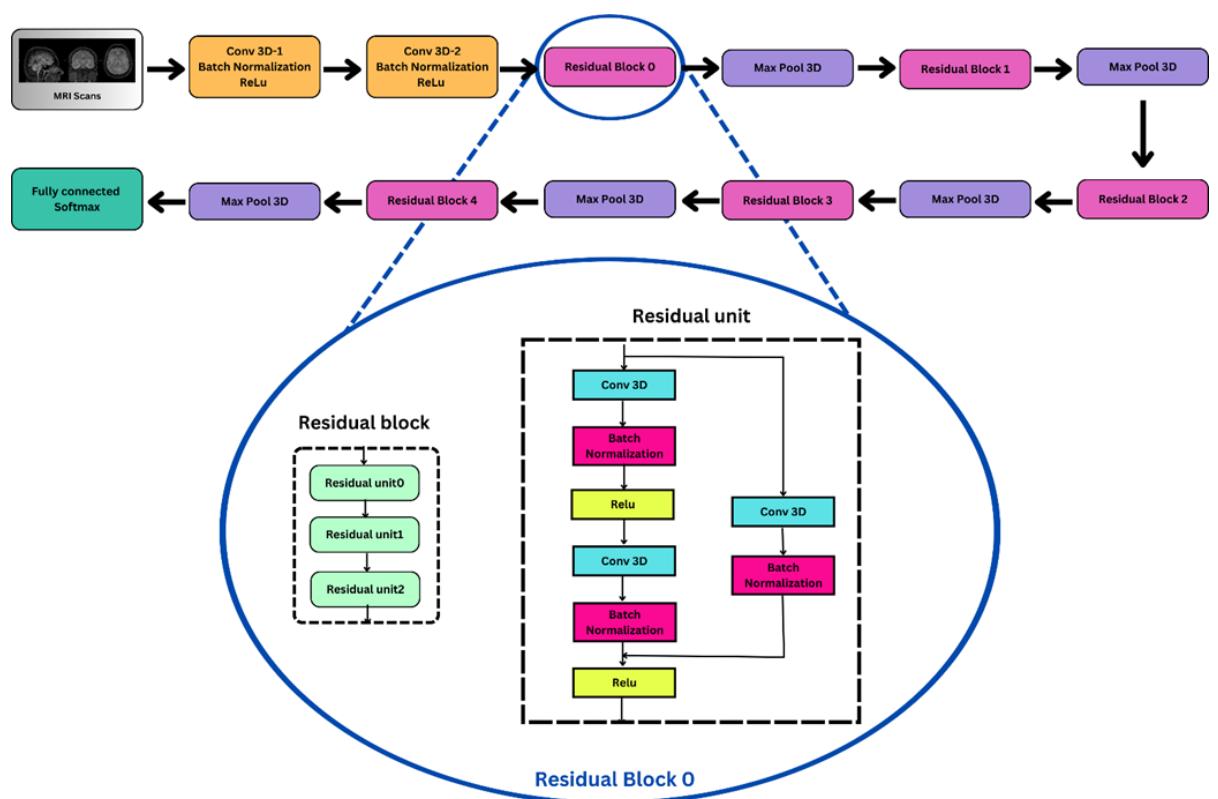
$$f(y) = \begin{cases} 0, & \text{if } y < 0 \\ y, & \text{if } y \geq 0 \end{cases} \quad (4.1)$$

Together, these two architectures significantly enhance the hybrid model's effectiveness: the 3D-CNN directly addresses the limitations in volumetric data handling, while the improved 3D-ResNet ensures efficient and stable learning through deep network layers. The combination of these architectures leverages their individual strength, effectively improving overall classification accuracy and robustness in early PD detection.

#### 4.4 Feature Extraction

Feature extraction is crucial for identifying relevant patterns and characteristics from MRI data to enhance predictive accuracy in diagnosing PD. In this study, specific layers from the deep

learning models were strategically chosen for feature extraction based on their ability to capture meaningful and discriminative information. From the custom-designed 3D-CNN, 2000 features were extracted from the fully connected layers (21st and 22nd layers), providing diverse high-level representations of the MRI data. Additionally, feature concatenation was employed at the 23rd layer, intelligently merging distinct sets of features into a single representative vector by selecting features with maximum discriminative values. Similarly, the 14th layer of the improved 3D-ResNet architecture was chosen for extracting another set of 2000 robust features. Consequently, a comprehensive feature set comprising 2000 features from each DL model was generated, ensuring a rich and informative dataset for subsequent fusion and optimization processes.



**Figure 4.3** Improved 3D-ResNet

## 4.5 Feature Fusion Techniques

Feature fusion involves integrating features from multiple deep neural network architectures to leverage their complementary strengths, thereby enhancing the overall effectiveness of the ML model. In this research, CCA was specifically chosen due to its robust ability to identify and exploit correlations between different feature sets, effectively combining the extracted features from the 3D-CNN and 3D-ResNet architectures.

### 4.5.1 Feature Concatenation

Feature concatenation merges distinct feature sets into a unified vector by selecting features with the highest discriminative power [64], as represented in Eq. 4.2. Concatenation of features from the 21st and 22nd layers of the constructed 3D-CNN significantly enhanced the accuracy of PD classification. The resulting composite vector, denoted as  $x_3$ , is derived by choosing the maximum values from the two feature vectors  $x_1 \in FV_1$  and  $x_2 \in FV_2$ . In cases where both vectors share identical elements,  $x_3$  selects a single instance, effectively eliminating redundancy. Table 4.2 outlines the pseudocode for the feature concatenation process. The final concatenated feature vector  $x_3$  dimensions are based on the feature vectors extracted, specifically  $FV_1 \times 1000$  and  $FV_2 \times 1000$ , as formulated in Eq. 4.2

$$\tilde{x}_3 = x_3 \leftarrow \text{maximum } (x_1, x_2) \text{ & } x_3 \in \text{Not Repeated} \quad (4.2)$$

**Table 4.2** Algorithm for Feature concatenation

---

#### Pseudocode for Feature concatenation

---

1. Input:

$x_1, x_2$ : Feature vector from space  $FV_1$  and  $FV_2$

$w_1, w_2$ : Weight for  $x_1$  and  $x_2$

2. Normalization:

Normalize ( $x_1$  and  $x_2$ )

3. Initialization:

$x_3$  = empty vector for the length ( $x_1$ )

used features = empty set

4. Feature Concatenation Loop:

for each index j from 0 to length ( $x_1$ ) - 1:

Calculate weighted  $x_1$  value =  $w_1 * x_1[j]$

Calculate weighted  $x_2$  value =  $w_2 * x_2[j]$

if weighted  $x_1$  value > weighted  $x_2$  value and  $x_1[j]$  not previously used:

Set  $x_3[j] = x_1[j]$

Else if weighted  $x_2$  value  $\geq$  weighted  $x_1$  value and  $x_2[j]$  not previously used:

Set  $x_3[j] = x_2[j]$

else:

Set  $x_3[j]$  to fallback logic (specific backup selection criteria)

Upload  $x_3[j]$  to use features

5. Output:

$x_3$ : Final concatenated feature vector with unique, highly discriminative elements

---

#### 4.5.2 Canonical Correlation Analysis (CCA)

In this study, CCA is applied to explore the relationships between two distinct sets of extracted features. A total of 2000 features were obtained—1000 each from the 3D ResNet and 3D CNN models. CCA serves as a statistical method to examine the associations between these two feature groups. Let us consider two matrices,  $A \in R^{pxn}$  and  $B \in R^{qxn}$ , where each matrix contains  $n$  training samples derived from different network architectures. This setup represents  $n$  observations, with a combined total of  $(p + q)$  features extracted per instance. Let  $S_{AA} = R^{pxp}$  and  $S_{BB} = R^{q \times q}$  it represents, the set covariance matrices of  $A$  and  $B$ , and  $S_{AB} = R^{pxq}$  denote the between-set covariance matrix, where  $S_{BA} = S_{AB}^T$ . The full covariance matrix, denoted as  $S \in R^{(p+q) \times (p+q)}$ , captures the complete statistical relationships between all pairs of features from both sets. This matrix, presented in Eq. 4.3, serves as the foundation for analyzing interdependencies across the combined feature space.

$$S = \begin{pmatrix} cov(A) & cov(A, B) \\ cov(B, A) & cov(B) \end{pmatrix} = \begin{pmatrix} S_{AA} & S_{AB} \\ S_{BA} & S_{BB} \end{pmatrix} \quad (4.3)$$

In this study, Canonical Correlation Analysis (CCA) is utilized for feature fusion due to its ability to compute optimal linear projections. These projections are represented as  $A^* = W_A^T A$  and  $B^* = W_B^T B$ , where the goal is to maximize the correlation between the transformed feature sets. This process is formally illustrated in Eq 4.4.

$$corr(A^*, B^*) = \frac{cov(A^*, B^*)}{var(A^*) \cdot var(B^*)} \quad (4.4)$$

where  $cov(A^*, B^*) = W_A^T S_{AB} W_B$ ,  $var(A^*) = W_A^T S_{AA} W_A$  and  $var(B^*) = W_B^T S_{BB} W_B$ . The optimization process uses Lagrange multipliers to maximize the covariance between the transformed feature sets  $A^*$  and  $B^*$  subject to the constraint that their variances remain equal to one, i.e.,  $var(A^*) = var(B^*) = 1$ . The resulting transformation matrices  $W_A$  and  $W_B$  are represented using Eq. 4.5,

$$\begin{cases} S_{AA}^{-1} S_{AB} S_{BB}^{-1} S_{BA} \hat{w}_A = \Lambda^2 \hat{w}_A \\ S_{BB}^{-1} S_{BA} S_{AA}^{-1} S_{AB} \hat{w}_B = \Lambda^2 \hat{w}_B \end{cases} \quad (4.5)$$

In this context,  $\Lambda^2$  represents the diagonal matrix containing the eigenvalues, which correspond to the squared canonical correlations, while  $\hat{w}_A$  and  $\hat{w}_B$  denote the associated eigenvectors. The eigenvectors linked to the non-zero eigenvalues are used to construct the transformation matrices. These transformed feature vectors are then either concatenated or summed to generate the final fused representation, denoted as  $Z$  enabling feature-level fusion. Here  $Z1$  and  $Z2$  is the Canonical Correlation Discriminant Features (CCDFs) mentioned in Eq. 4.6 and 4.7. The fused vector ( $Z$ ) is applied to the optimization-based feature elimination module to refine and select essential features that contribute significantly to the model's predictive accuracy.

$$Z_1 = \begin{pmatrix} A^* \\ B^* \end{pmatrix} = \begin{pmatrix} W_A^T A \\ W_B^T B \end{pmatrix} = \begin{pmatrix} W_A & 0 \\ 0 & W_B \end{pmatrix}^T \begin{pmatrix} A \\ B \end{pmatrix} \quad (4.6)$$

$$Z_2 = A^* + B^* = W_A^T A + W_B^T B = \begin{pmatrix} W_A \\ W_B \end{pmatrix}^T \begin{pmatrix} A \\ B \end{pmatrix} \quad (4.7)$$

## 4.6 Feature optimization

The main objective of feature optimization is to pinpoint the most impactful features that contribute to accurate predictions, while also reducing the system's computational burden. By selecting only the top-ranked features, the dimensionality of the dataset is effectively minimized. This not only simplifies data storage and management but also makes the training and inference processes more efficient and manageable.

### 4.6.1 Mathematical model of WOA

The Whale Optimization Algorithm (WOA), introduced by Mirjalili and Lewis, draws inspiration from the unique hunting technique of humpback whales known as bubble-net feeding [108], [109]. In nature, whales use spiral patterns of bubbles to encircle and trap their prey. This behavior is mimicked in WOA to efficiently navigate and solve complex optimization problems by simulating the whales' search and trapping strategies.

**Encircling Prey:** Humpback whales hunt by identifying and surrounding their prey. WOA simulates this behavior by treating the current best solution as an approximation of the optimal one. Once this leading solution (or search agent) is determined, the other agents begin to move toward it, effectively imitating the whales' encircling action. This mechanism is captured mathematically in Eq. 4.8 and 4.9. Here,  $\vec{Z}(u)$  represents the agent's position,  $u$  represents the iteration and  $\vec{Z}^*$  represents the optimal solution. In Eq. 4.10 and 4.11  $\vec{A}$  and  $\vec{D}$  indicate convergence values. The random number is  $\vec{s}[0,1]$ , and  $\vec{s}$  stands for the vector that decreases linearly from 2 to 0 during an iteration.

$$\vec{Z}(u+1) = |\vec{Z}^*(u) - \vec{A} \cdot \vec{F}|, \quad (4.8)$$

$$\vec{F} = |\vec{D} \cdot \vec{Z}^*(u) - \vec{Z}(u)|, \quad (4.9)$$

$$\vec{A} = 2 \cdot \vec{a} \cdot \vec{s} - \vec{a}, \quad (4.10)$$

$$\vec{D} = 2 \cdot \vec{s} \quad (4.11)$$

**Bubble-net attacking Strategy:** Humpback whales use a distinctive hunting method that involves spiraling and gradually closing in on their prey. In the Whale Optimization Algorithm (WOA), this behavior is replicated by gradually reducing the parameter  $\vec{a}$  as shown in Eq. 4.11, which effectively narrows the search space and intensifies the focus around promising solutions. Since  $\vec{A}$  is dependent on  $\vec{a}$ , it also decreases linearly from 2 to 0. The spiral motion, a key aspect of the whale's attack, is mathematically captured in Eq. 4.12 and 4.13, representing how the algorithm mimics this dynamic behavior to refine its search process.

$$\vec{F}^* = |\vec{Z}^*(u) - Z(u)|, \quad (4.12)$$

$$\vec{Z}(u+1) = e^{bk} \cdot \cos(2\pi k) \cdot \vec{F}^* + \vec{Z}^*(u) \quad (4.13)$$

The variable  $F^*$  represents the distance between the whale and its target. In this context,  $b$  is a constant that defines the shape of the logarithmic spiral, while  $k$  is a randomly chosen value within the range  $[-1, 1]$ . In nature, humpback whales alternate between two strategies—spiralizing toward the prey or directly closing in—and this behavior is reflected in WOA. The algorithm assigns a 50% probability to each strategy, determined by the parameter  $p$  in Eq. 4.14, which is randomly selected between  $[0, 1]$ .

$$\vec{Z}(u+1) = \begin{cases} \vec{Z}^* - \vec{A} \cdot \vec{F}, & \text{if } p < 0.5 \\ e^{bk} \cdot \cos(2\pi k) \cdot \vec{F}^* + \vec{Z}^*(u), & \text{if } p \geq 0.5 \end{cases} \quad (4.14)$$

In the exploration phase of WOA,  $\vec{A}$  is randomly assigned a value between  $[-1, 1]$ , prompting the search agents to diverge from the reference whale. This promotes a global search as the updated position of a search agent is determined by randomly selecting another agent. The mathematical formulation of this exploration mechanism is outlined in Eq. 4.15 and 4.16, where

$\overrightarrow{Z_{rand}}$  is a random location for the random whale that is chosen from the current population.

$$\vec{Z}(u+1) = \overrightarrow{Z_{rand}} - \vec{A} \cdot \vec{F}, \quad (4.15)$$

$$\vec{F} = |\vec{C} \cdot \overrightarrow{Z_{rand}} - \vec{Z}| \quad (4.16)$$

#### 4.6.2 WOA for feature selection

In this study, the Whale Optimization Algorithm (WOA), inspired by the natural foraging behavior of humpback whales, is employed to improve feature selection for Parkinson's Disease classification. A total of 2000 features were extracted—half from the FC-3 layer of a 3D-CNN and the other half from the FC-4 layer of a 3D ResNet network. Effective feature selection plays a vital role in refining data representation, enhancing model accuracy, and reducing unnecessary complexity. Selecting an optimal subset of features not only boosts classification performance but also leads to faster learning, better generalization, and reduced computational demands. However, challenges such as getting trapped in local optima or dealing with high computational costs often arise. To overcome these, a robust global search strategy like WOA is essential. In the context of WOA, each candidate solution represents a possible subset of features analogous to a whale's position in the search space. The total number of features that can be selected corresponds to the number of features in the original dataset, denoted as  $N$ . Each solution is evaluated based on two objectives: the number of features selected and the classification accuracy achieved. Ideally, a good solution selects fewer features while still maintaining high accuracy. To assess the quality of each feature subset, a fitness function is used that combines both criteria minimizing the number of features and maximizing classification accuracy. The accuracy is computed using a K-Nearest Neighbors (KNN) classifier. The representation of each solution is critical in metaheuristic design. Here, solutions are encoded as binary vectors of length  $N$ , where each element is either "1" or "0". A "1" indicates that a feature is selected, while a "0" means it is excluded. This binary encoding offers an efficient way to manage feature inclusion within each candidate solution. The final fitness function, defined in Eq. 4.17, is carefully constructed to maintain a balance between reducing

the feature count and maximizing model performance.

$$f_{\theta} = \alpha \cdot \gamma_R(D) + (1 - \alpha) \frac{|R|}{|N|} \quad (4.17)$$

Here,  $f_{\theta}$  is the fitness function used to evaluate the quality of each solution,  $\gamma_R(D)$  represents the classification error rate calculated using the KNN classifier.  $|R|$  refers to the number of features selected (i.e., the cardinality of the subset), and  $|N|$  indicates the total number of features in the original dataset. The parameter  $\alpha$  and  $\beta = (1 - \alpha)$  are introduced to balance the trade-off between classification accuracy and feature subset size. Specifically,  $\alpha$  controls the weight assigned to accuracy, while  $\beta$  governs the influence of feature reduction. This design ensures a practical and fair evaluation of candidate solutions, as outlined in [110].

The optimization begins by randomly initializing the coefficient vectors  $\vec{A}$  and  $\vec{D}$ , along with an initial best solution vector  $\vec{Z}^*$ , randomly selected from the search space. The WOA then iteratively updates the position of each whale based on Eq. 4.8 and 4.9, representing the encircling and exploration behaviors. In each iteration, the fitness of all candidate solutions is computed using the defined fitness function. If a better solution is found, the current best solution vector  $\vec{Z}^*$  is updated accordingly. This process continues for a pre-set number of iterations and is repeated across multiple independent runs to ensure consistency and robustness. The optimization concludes when the maximum number of iterations is reached. A visual representation of the WOA-based feature selection process is illustrated in Figure 4.4.

## 4.7 Results and Analysis

This section presents and interprets the experimental outcomes obtained from the proposed framework for PD classification using MRI neuroimaging data. The study focuses on the effective differentiation of PD, prodromal, and control subjects through DL models trained on pre-processed T2-weighted MRI scans. The proposed methodology incorporates multiple DL architectures, a custom architecture 3D-CNN and an improved 3D-ResNet model, tailored to

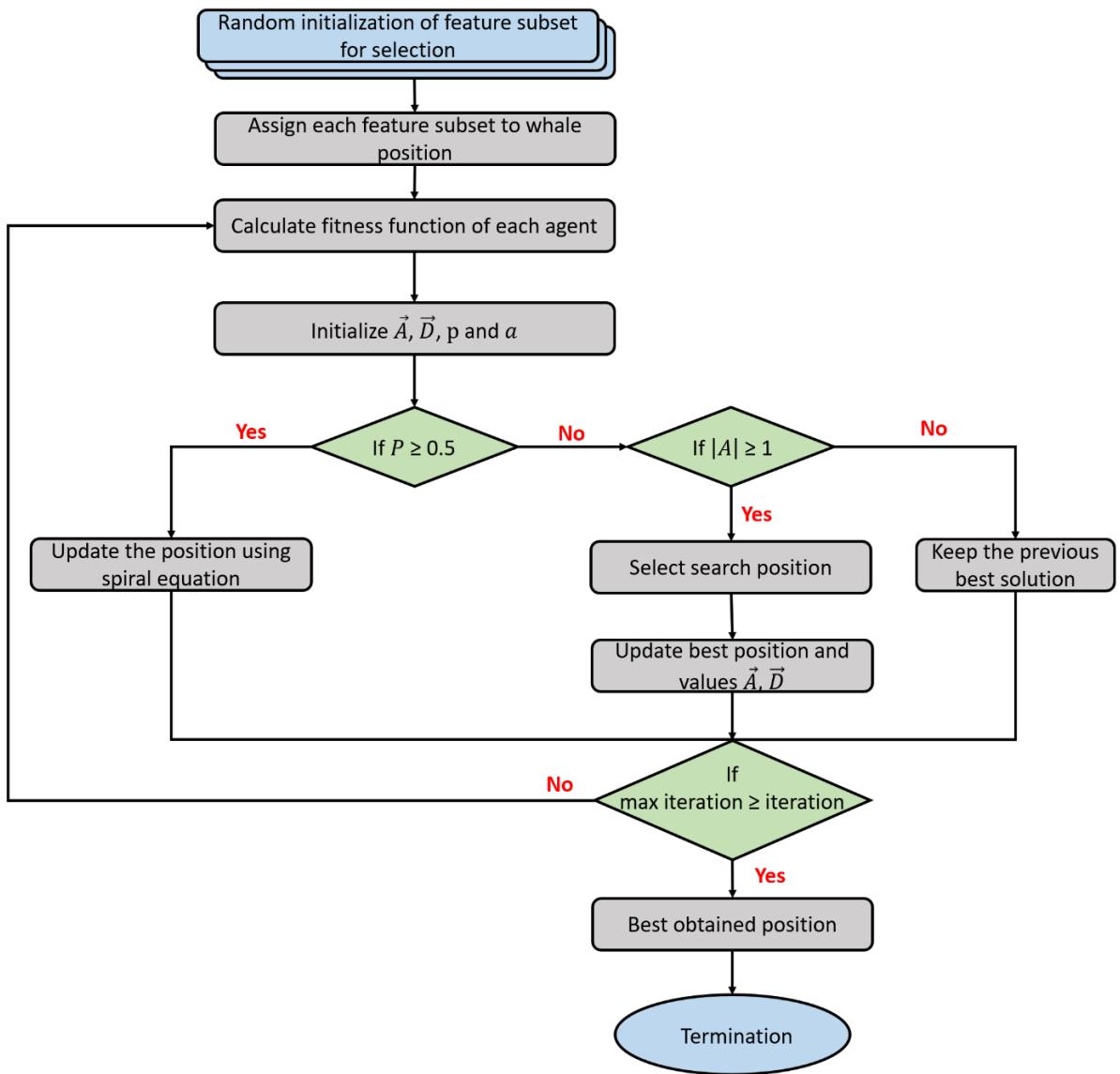
extract spatial and contextual information from volumetric brain data. This study explores the impact of optimisation techniques on classification accuracy. Two experimental setups were compared: one without any optimisation and another employing a WOA following feature-level fusion via CCA.

#### **4.7.1 Model-1 (Proposed 3D-CNN) architecture analysis**

A total of 303 MRI scans were sourced from the PPMI (Parkinson's Progression Markers Initiative) database, representing three distinct subject groups. These scans, originally in the DICOM (Digital Imaging and Communication in Medicine) format, consisted of 100 to 200 slices per subject. For consistency and ease of processing, each subject's DICOM data was converted into a single 3D NIfTI file. The NIfTI images were then pre-processed using skull stripping and bias field correction techniques to improve overall image quality. After preprocessing, all volumes were resized to a standardized dimension of **56×56×56** for further analysis.

An initial investigation was carried out using a baseline 3D Convolutional Neural Network (3D-CNN), detailed in Table 4.3, followed by an ablation study to identify the most effective network configuration. The ablation was conducted in multiple stages, progressively modifying the network architecture to improve classification performance.

- **Stage 1 (Base Model):** An 8-layer 3D-CNN was built, including an input layer, two 3D convolutional layers, a ReLU activation layer, one pooling layer, batch normalization, a flattening layer, and a final dense SoftMax classification layer. This configuration achieved an accuracy of **82.02%**.
- **Stage 2 (Variant 1):** The architecture was extended to 9 layers by adding an additional 3D convolutional layer before the activation and pooling stages. This slight enhancement improved the model's accuracy to **85.75%**.



**Figure 4.4** Feature Selection process using WOA

- **Stage 3 (Variant 2):** A more complex 17-layer design was implemented, consisting of an input layer followed by two initial convolutional blocks, ReLU activations, pooling and batch normalization, and then another block of three convolutional layers with activations, pooling, and normalization, before the final fully connected and SoftMax layers. This version achieved **88.76%** accuracy.

- **Stage 4 (Variant 3):** The most refined architecture with 24 layers was developed. It started with two 3D convolutional layers, followed by activations, pooling, and normalization. This was followed by three additional convolutional layers with their respective activations and pooling, a second normalization block, and then two more convolutional layers with activations, pooling, and batch normalization. The final part consisted of a flatten layer and a SoftMax classifier. This configuration delivered the highest performance, reaching an accuracy of **89.43%**.

**Table 4.3** Ablation Study of the Constructed 3D-CNN Architecture

**Study 1:** Modify the convolutional layer and max pool layer

Model	No. of Conv_layer	No. of pooling layer	Test accuracy (in %)	Finding
Base Model	2	1	82.02	Lowest accuracy
Variant 1	3	1	85.75	Intermediate
Variant 2	2+3	1+1	88.76	Intermediate
Variant 3	<b>2+3+2</b>	<b>1+1+1</b>	<b>89.43</b>	<b>Highest</b>

**Study 2:** Modifying the different parameters for four approaches

Parameters	Base Model	Variant 1	Variant 2	Variant 3
Pooling layer				
Max	✓		✓	✓
Average		✓		
Activation function				
ReLU	✓	✓	✓	✓
PreLu				
Batch Size				
32			✓	✓
64	✓		✓	
Flatten layer				
Flatten	✓	✓	✓	✓
Global max				
Optimizer				
Adam		✓	✓	✓
SGD	✓			
Learning rate				
0.01	✓			
0.001		✓	✓	✓
Epoch	30	30	30	30
Evaluation metrics				

Table 4.3 (Continues)

Accuracy	84.96	88.25	90.28	93.41
Precision	87.34	89.93	91.83	95.76
Recall	81.54	84.28	87.72	88.98
F1-score	85.32	87.43	90.81	91.68

Building upon the evaluation of the base 3D-CNN model and its three architectural variants, the next phase of this study involved fine-tuning the models through comprehensive hyperparameter optimization. As outlined in Table 4.3 under Study 2, various hyperparameters were systematically varied to determine the optimal configuration that maximizes classification accuracy. Parameters explored included the type of pooling (e.g., max or average), activation functions, batch sizes, flattening methods, optimizers, and learning rates. Each of these was tested across the base model and all three variants to identify the best-performing setup. Among all configurations, Variant 3 demonstrated superior performance, outperforming the other models across multiple evaluation metrics. The ablation process proved highly effective boosting classification accuracy from 89.43% to 93.41%, marking a significant improvement of 3.98%. Such enhancement is particularly valuable in medical diagnosis tasks, where precision is critical. In this study, model evaluation followed standard medical diagnostic metrics, including accuracy, precision (positive predictive value or PPV), recall (sensitivity or true positive rate), and the F1-score.

The analysis was divided into three stages:

- **Analysis 1** focused on features extracted from the FC-1 (21st layer) of the 3D-CNN. Multiple machine learning classifiers were applied, and the Random Forest (RF) classifier delivered the highest accuracy at 88.7%, outperforming the rest.
- **Analysis 2** used features from the FC-2 (22nd layer) of the network. This time, the Gradient Boosting (GB) classifier achieved the best result, with an accuracy of 90.1%.
- **Analysis 3** involved combining the features from both FC-1 and FC-2 layers into a unified

representation termed FC-3. These combined features were then fed into various ML classifiers. This fusion approach led to a further boost in performance, with the Gradient Boosting model reaching a peak accuracy of 93.4%, surpassing all previous configurations.

The consistent use of precision, recall, accuracy, and F1-score across all analyses provided a clear and balanced view of model effectiveness. Overall, these results underscore the value of carefully selecting feature extraction layers and optimizing hyperparameters to enhance the prediction of PD.

#### **4.7.2 Model-2 (Improved 3D-ResNet)**

The detailed model architecture of improved 3D-ResNet is shown in Figure 4.3. Features were extracted from a fully connected layer, FC-4, and then processed with a focus on classification using various ML models. The Gradient boosting-based ML algorithm achieved a maximum accuracy of 90%. The detailed performance analysis of the proposed framework is compared with existing pre-trained models and is elaborated in Table 4.4.

**Table 4.4** Comparative analysis of PD prediction performance using different architectural models and Feature Extraction Methods

<b>3DCNN</b>				
<b>Analysis 1:</b> PD prediction using extracted features from 3D-CNN FC-1 layer				
Methods	Accuracy	Recall	Precision	F1-score
SVM	85.6	82.4	88.5	86.3
KNN	84.2	82.2	86.2	83.7
GB	87.1	83	88.5	87.2
RF	88.7	85.2	89.3	87.6
<b>Analysis 2:</b> PD prediction using extracted features from 3D-CNN FC-2 layer				
Methods	Accuracy	Recall	Precision	F1-score
SVM	85.9	84	86.4	84.3

Table 4.4 (Continued)

KNN	86.2	82.2	87.2	83.9
GB	90.1	86	92	88.9
RF	88.3	85.9	89.3	87.6
<b>Analysis 3:</b> PD prediction using feature fusion from FC-1 and FC-2 layer (i.e.FC-3)				
Methods	Accuracy	Recall	Precision	F1-score
SVM	90.5	88	91.9	89.1
KNN	88.9	86.2	90.2	87.4
GB	<b>93.5</b>	91	96.2	92.5
RF	91.8	87.4	93.3	90.3
<b>Improved 3DResNet</b>				
Methods	Accuracy	Recall	Precision	F1-score
SVM	85.3	82.1	86.3	85.1
KNN	88.1	83.2	89.9	87.4
GB	90	88.3	92.5	91
RF	87.2	85.7	89.1	86.2

#### 4.7.3 Model-3 (Fusion without Optimization)

In this model, features were collected from the FC-3 layer of a 3D-CNN and the FC-4 layer of a 3D ResNet. These feature sets, originating from two different deep learning architectures, were combined using CCA. The goal of this fusion strategy was to merge the complementary information captured by each model, thereby improving the overall classification performance. Following the fusion, the GB classifier identified as the top-performing model in earlier evaluations was applied to the fused features. To further validate and enhance its reliability, a 15-fold cross-validation was conducted. This evaluation led to an additional 4.8% increase in classification accuracy, demonstrating the effectiveness of the fusion approach even without feature optimization.

**Table 4.5** Performance analysis of models 3 and 4

Model 3 (without optimization)					
S.no	Cross-validation	Accuracy	Recall	Precision	F1-Score
1	5-fold	0.916	0.886	0.935	0.905
2	10-fold	0.924	0.882	0.928	0.918
3	15-fold	0.948	0.904	0.956	0.938
Model 3 (with optimization)					
S.no	Cross-validation	Accuracy	Recall	Precision	F1-Score
1	5-fold	0.959	0.907	0.986	0.929
2	10-fold	0.962	0.893	0.988	0.934
3	15-fold	0.972	0.906	0.985	0.931

#### 4.7.4 Model-4 (Feature Fusion with WOA based optimization)

While feature fusion using CCA effectively combines information from multiple models, it often results in high-dimensional feature spaces and increased sensitivity to noise or outliers. To address these challenges, this study introduces a bio-inspired optimization strategy WOA to refine the fused feature set. The WOA-based feature selection method helps identify the most relevant and discriminative features from the combined output of the FC-3 layer of the 3D-CNN and the FC-4 layer of the 3D ResNet. This optimization enhances the quality of feature representation and boosts the classifier's ability to generalize. Using the GB classifier along with 15-fold cross-validation, the optimized feature set achieved an impressive accuracy of 97.2%. This significant improvement over the unoptimized model highlights the effectiveness of WOA in reducing redundancy and selecting the most impactful features. A detailed comparison of Model-3 and Model-4 performance is provided in Table 4.5.

## 4.8 Discussion

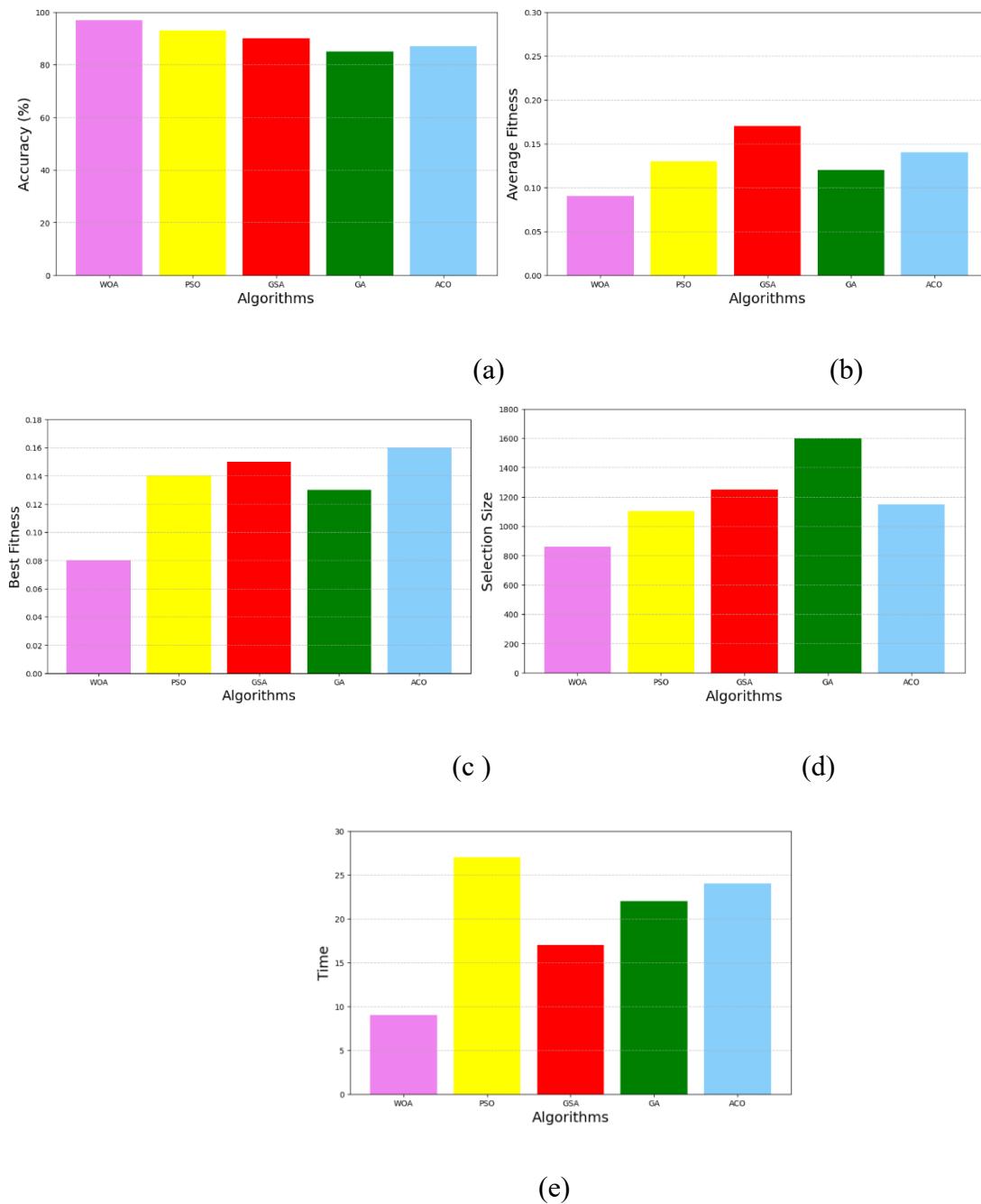
This section presents a comparative analysis of four distinct approaches implemented for PD diagnosis. The study began with the development of Model-1, a custom 3D-CNN architecture, which was iteratively enhanced through ablation studies. This deep learning model demonstrated strong automatic feature extraction capabilities, achieving notable performance: 93.5% accuracy, 91.0% precision, 96.2% recall, and an F1-score of 92.5%. These results highlight the effectiveness of end-to-end 3D-CNNs in analyzing T2-weighted brain MRI scans for PD classification. In comparison, Model-2 utilized an enhanced 3D-ResNet for feature extraction, followed by classification using multiple machine learning algorithms. Although its accuracy of 90% was slightly lower than Model-1, it still proved to be a competent framework for identifying early-stage PD, aligning with earlier findings in the literature [21].

To further enhance performance, Model-3 introduced a feature fusion strategy by combining outputs from both the 3D-CNN and 3D-ResNet models using CCA. This fusion approach capitalized on the complementary strengths of both architectures and led to an approximate 2% improvement in accuracy over Model-1. However, the added dimensionality from fusion increased the risk of overfitting and introduced additional computational overhead. To address these limitations, Model-4 incorporated a feature optimization stage using the Whale Optimization Algorithm (WOA). This bio-inspired method was applied to reduce redundancy in the fused feature set and enhance classification robustness. For a fair comparison, WOA was benchmarked against four other well-known metaheuristic algorithms Particle Swarm Optimization (PSO), Gravitational Search Algorithm (GSA), Genetic Algorithm (GA), and Ant Colony Optimization (ACO) all recognized for their effectiveness in solving feature selection and search space optimization problems. The key parameter settings for each algorithm are summarized in Table 4.6.

**Table 4.6** Parameter values for all algorithm

Algorithms	Parameter	Value
PSO	Cognitive factor (c1)	2
	Social factor (c2)	2
	Inertia weight (w)	0.9
GSA	Gravitational constant(G0)	100
	Constant (alpha)	20
GA	crossover rate (CR)	0.8
	Mutation rate (MR)	0.01
	Tournament size (Ts)	3
ACO	Pheromone value (tau)	1
	Heuristic desirability (eta)	1
	Control pheromone (alpha)	1
	Control heuristic (beta)	0.1
	Pheromone trail decay coefficient (rho)	0.2
WOA	Constant (b)	1
	Threshold(thres)	0.5
Search settings	Population Size	100
	Number of iterations	200
	Lower and upper bound	0 and 1
	No. of independent runs	20

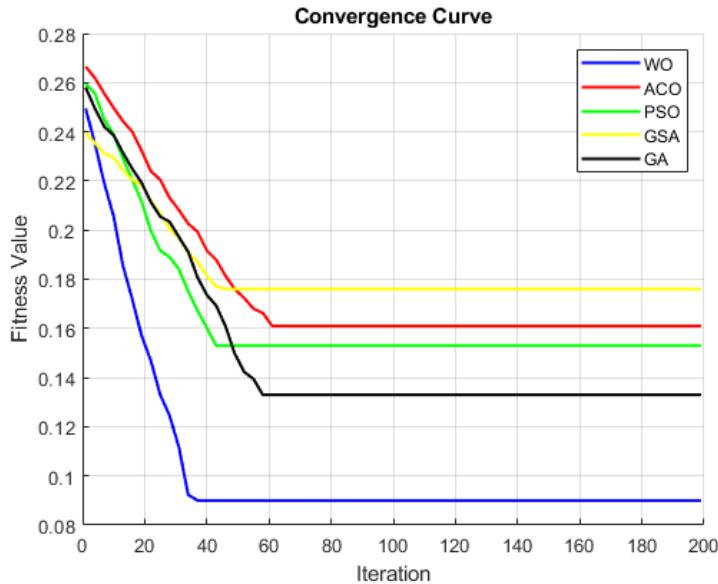
The effectiveness of the optimization algorithms was evaluated using multiple performance metrics, including classification accuracy, best and average fitness scores, number of selected features, and computational time. A detailed comparison of these metrics is presented in Figure 4.5. Among the algorithms tested, the WOA consistently outperformed the others in terms of classification accuracy. Additionally, WOA showed favorable results across other evaluation criteria, including reduced feature subset size and competitive computational efficiency. These advantages suggest that WOA not only enhances predictive performance but also maintains a manageable model complexity. Overall, the comparative analysis highlights that WOA delivers superior optimization performance when compared to other metaheuristic approaches, making it a strong candidate for feature selection in medical imaging-based PD diagnosis.



**Figure 4.5** The overall performance comparison of algorithms across different metrics (a) Accuracy (b) Average Fitness (c) Best Fitness (d) Selection size (e) Time

Figure 4.6 presents the convergence curves for various optimization algorithms applied to the PPMI dataset. These curves reflect the fitness values achieved over 200 iterations, providing insight into the optimisation behaviour of each metaheuristic method. Among the algorithms

compared, the WOA consistently reached lower fitness values than GA and PSO, indicating superior performance in finding optimal feature subsets.



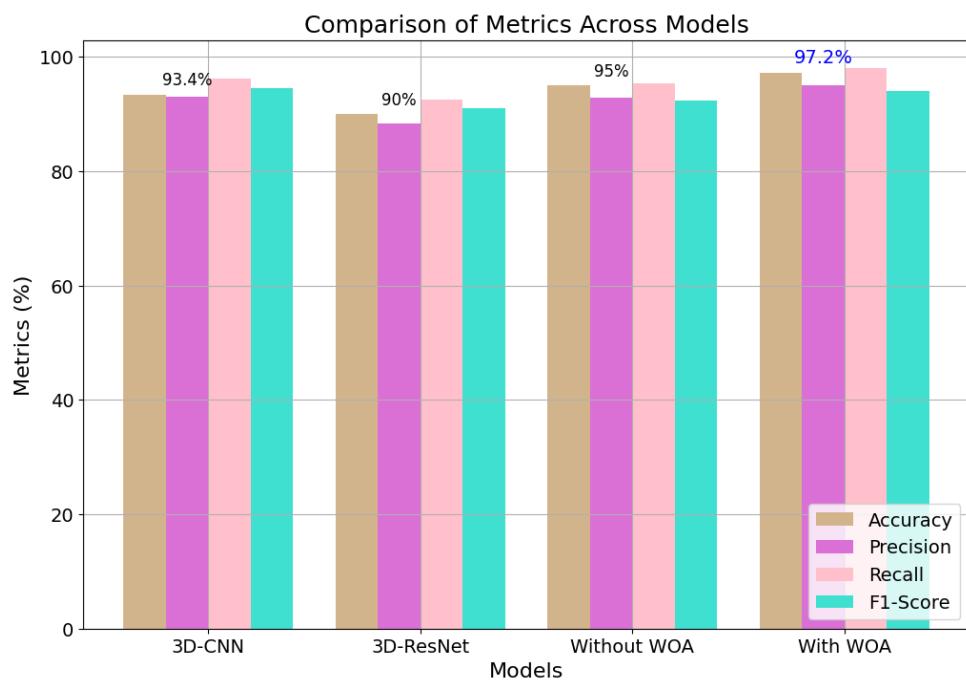
**Figure 4.6** Convergence curves generated using various algorithms for the PPMI dataset.

The integration of WOA in Model-4 led to a significant 5% improvement in accuracy compared to the baseline Model-1. This enhancement translated into strong overall classification performance, with the model achieving an accuracy of 97.2%, precision of 90.6%, recall of 98.5%, and an F1-score of 92.1%. These results emphasize the effectiveness of incorporating an optimized feature selection strategy. The comparative performance of all models developed in this study is visually summarized in Figure 4.7.

#### 4.9 Summary

This chapter presented a comprehensive and robust hybrid DL framework to improve the accuracy and early detection capability of PD using whole-brain MRI data. Key highlights and significant findings from this work include:

- Successful development and validation of custom-designed 3D-CNN and enhanced 3D-ResNet architectures, rigorously optimized via hyperparameter tuning and ablation studies.
- Effective integration of spatially complementary features from both deep-learning models using CCA.
- Implementation of WOA for feature optimization, significantly reducing computational complexity.
- Superior classification performance, achieving an exceptional accuracy of 97.2%, significantly outperforming traditional ML methods and individual CNN models.



**Figure 4.7** Comparison of performance metrics for the proposed method

- Demonstration of the proposed model's generalizability and clinical applicability, addressing the critical need for reliable early PD diagnosis.

Collectively, these results underscore the efficacy of combining advanced DL architectures with intelligent optimization strategies, presenting substantial potential for enhancing diagnostic accuracy and clinical outcomes in PD.

## **CHAPTER 5: MULTIMODAL AI FRAMEWORK FOR PERSONALISED PD DIAGNOSIS**

### **5.1 Introduction**

PD is a complex neurodegenerative disorder marked by a range of motor impairments such as tremors, bradykinesia, and rigidity, as well as non-motor symptoms including cognitive decline and emotional disturbances. The absence of a definitive cure emphasizes the critical need for early detection and continuous monitoring, both of which are essential for initiating timely interventions that can slow disease progression and improve quality of life [9], [111]. However, early diagnosis remains a significant clinical challenge due to the overlapping symptomatology with other neurological conditions and the subjective nature of conventional diagnostic methods. Currently, the diagnosis of PD predominantly relies on clinical assessments by neurologists, supported by neuroimaging findings and limited biochemical testing [112]. These approaches, although valuable, are often constrained by inconsistent sensitivity to early symptoms, limited integration of heterogeneous data sources, and an overall lack of interpretability in automated decision-making tools [91]. Furthermore, in low-resource environments such as rural or remote regions of India, access to specialists, advanced imaging modalities, and follow-up care is considerably limited, further complicating timely and accurate diagnosis [14].

To address these gaps, this chapter presents a novel multimodal artificial intelligence (AI) framework designed to provide a unified, explainable, and accessible diagnostic solution for PD. The core motivation stems from the need to systematically integrate diverse clinical, neuroimaging, and biochemical data to develop a diagnostic tool that is not only accurate but also interpretable and deployable in real-world settings. Unlike traditional AI models that often function as opaque black boxes, the proposed system emphasises transparency through the incorporation of explainable AI (XAI) and personalised decision support mechanisms [39],

[113], [114]. Importantly, the framework aims to bridge the disparity between cutting-edge computational models and their practical clinical application, especially in under-resourced healthcare environments.

The proposed system leverages a hybrid architecture that combines a 1d Convolutional Neural Network (1d-CNN) for structured classification with a fine-tuned Large Language Model (LLM) based on Chatgpt-4.0 to enable real-time interaction and personalised diagnostics. The 1d-CNN is trained on multimodal data derived from the Parkinson's Progression Marker Initiative (PPMI) database, which includes radiomics features from MRI scans, specific binding ratios (SBR) from dopamine transporter (DaT) SPECT imaging, cerebrospinal fluid (CSF) protein biomarkers, and standardized clinical scores such as the Unified Parkinson's Disease Rating Scale (UPDRS) and Montreal Cognitive Assessment (MoCA). The fusion of these modalities not only enhances diagnostic accuracy but also captures distinct aspects of disease progression [115]. Furthermore, the LLM module is fine-tuned to provide natural language explanations, generate personalised medical summaries, and respond to clinician and patient queries, thereby making the system interactive and human-centred [116], [117]. The entire framework is deployed via a secure, cloud-based platform that supports seamless data upload, inference, and diagnostic report generation. This enables patients and clinicians, regardless of geographical constraints, to access expert-level insights, making the solution highly scalable and practical for real-time healthcare delivery. The key contributions of this chapter are:

- The development of a multimodal diagnostic pipeline that integrates neuroimaging, biochemical, and clinical data for improved PD stratification.
- A customised 1d-CNN architecture capable of performing accurate multiclass classification among control, prodromal, and PD subjects.
- The incorporation of ratio-based engineered features to improve early-stage PD detection.
- The fine-tuning of a domain-specific LLM to provide explainable diagnostic summaries, interpret outputs, and support real-time clinical interaction.

- A fully functional cloud-based interface for secure data handling, report generation, and longitudinal health record management, with a focus on usability in low-resource clinical settings.

## 5.2 Data collection

To construct and evaluate the proposed multimodal diagnostic framework for PD, this study utilised publicly available data from the PPMI repository. The PPMI is a globally recognized, multicentre, and longitudinal observational study aimed at identifying reliable biomarkers for the early detection and monitoring of PD. Data access was obtained through the official PPMI portal [<https://www.ppmi-info.org/access-data-specimens/download-data>], adhering to the standard data use and ethical compliance protocols. A total of 150 participants were selected based on the availability and completeness of both imaging and clinical data. Each subject included in the cohort had undergone high-resolution T2-weighted MRI scanning and DaT imaging at their most recent clinical visit. In addition to imaging, the dataset encompassed comprehensive clinical evaluations, SBR values derived from the SPECT scans, and CSF biomarker measurements. The inclusion criteria ensured representation across the full spectrum of disease states, comprising individuals diagnosed with PD, prodromal subjects (preclinical stage), and healthy control (HC) participants, thus enabling robust comparative analysis and disease stratification. The compiled dataset consisted of four primary modalities: Anatomical neuroimaging (MRI) for radiomics-based structural brain analysis, Functional imaging (DaT-SPECT) for dopaminergic activity estimation, Biochemical markers (CSF proteins) for molecular-level characterisation, and Standardised clinical assessments including UPDRS and Moca scores for phenotypic profiling. By integration of these multimodal inputs, the framework aims to capture structural, functional, molecular, and behavioural signatures of PD within a unified pipeline.

### **5.3 Delineation of multimodal data features**

To improve diagnostic accuracy and capture the multidimensional nature of PD, this study leverages a comprehensive set of features derived from structural imaging, functional imaging, biochemical markers, and clinical assessments. Each modality was carefully selected based on its proven relevance in reflecting distinct pathological signatures of PD. The extracted features were employed both for training the 1d-CNN classifier and for guiding downstream reasoning in the fine-tuned LLM.

#### **5.3.1 Neuroimaging data with radiomics features**

T2-weighted MRI scans formed the backbone of structural analysis, offering high-resolution insights into brain morphology. All scans were acquired using a SIEMENS Prisma 3.0T system with standardised acquisition parameters (TR: 2300 ms, TE: 3.0 ms, TI: 900 ms, flip angle: 9°, isotropic resolution: 1 mm). To ensure compatibility with existing neuroimaging pipelines, Digital Imaging and Communications in Medicine (DICOM) images were converted into Neuroimaging Informatics Technology Initiative (NIfTI) format. Radiomics features were extracted across axial, sagittal, and coronal planes to capture subtle texture and intensity-based variations that may not be visible through qualitative radiological interpretation. These features provided a quantitative characterization of regional brain structures affected by neurodegeneration and were instrumental in distinguishing between control, prodromal, and PD groups.

#### **5.3.2 SBR values from SPECT Imaging**

DaT-SPECT imaging was employed to assess presynaptic dopaminergic function. Scans were performed using I-123 Ioflupane as the radiotracer and acquired via a SIEMENS system with step-and-shoot acquisition (3° angular increments across 180°). SBR were computed for four striatal regions: left/right caudate and left/right putamen. These values served as critical quantitative biomarkers, reflecting dopaminergic depletion, a key pathophysiological feature of

PD. Declining SBR values, particularly in the putamen, were associated with disease progression and motor symptomatology.

### **5.3.3 Biological features (CSF protein biomarkers)**

CSF analysis plays a crucial role in revealing the biochemical changes that accompany neurodegeneration in PD. Four protein biomarkers were extracted and included in the study:

- $\alpha$ -synuclein ( $\alpha$ -syn): A hallmark protein associated with PD pathology, known for its aggregation in Lewy bodies.
- Amyloid-beta 1–42 ( $A\beta_{1-42}$ ): A key indicator of amyloid plaque formation and neuroinflammatory processes.
- Total tau (t-tau): Reflects general neuronal injury and neurodegeneration.
- Phosphorylated tau (P-tau<sub>181</sub>): A marker specific to tau pathology and microtubule destabilization.

These biomarkers allowed the model to differentiate PD from other neurodegenerative conditions and were particularly valuable for characterizing prodromal subjects, where conventional imaging may not be sufficiently sensitive [104].

### **5.3.4 Clinical data**

To capture the comprehensive clinical profile of PD, both motor and non-motor symptoms were evaluated using standardised neurological assessment instruments. The primary tools employed in this study were the MDS-UPDRS and MoCA, each contributing unique insights into different dimensions of disease manifestation and severity [118]. The MDS-UPDRS comprises four subscales, each targeting a specific domain:

- Part I: Non-Motor Experiences of Daily Living (Range: 0–52) – Evaluates symptoms such as cognitive decline, mood alterations, behavioral changes, and sleep disturbances.

- Part II: Motor Experiences of Daily Living (Range: 0–52) – Assesses the impact of PD on everyday motor functions including speech, handwriting, and personal hygiene.
- Part III: Motor Examination (Range: 0–132) – A clinician-rated assessment that quantifies motor impairments such as tremors, rigidity, bradykinesia, posture, and gait.
- Part IV: Motor Complications (Range: 0–24) – Measures therapy-related complications, including dyskinesias and motor fluctuations.

Higher UPDRS scores reflect greater disease burden. MoCA was used to evaluate cognitive functions, with scores below 26 indicating potential impairment. These clinical measures were not only used as input features but also served as ground truth labels for model training and validation. Their integration with imaging and biochemical data enabled a holistic view of disease manifestation. Table 5.1 summarizes the demographic variables, clinical scores, SBR values, and CSF biomarkers across control, prodromal, and PD groups, including statistical significance levels based on ANOVA and chi-square tests.

**Table 5.1:** Details of the multimodal data with p-values of the biomarkers used in the current research

Variable	PD subjects	Prodromal subjects	Control Subjects	p-value
<b>Clinical data</b>				
Gender (F/M)	25/25	21/29	23/27	>0.05 <sup>b</sup>
Age				
History of pd (Yes/no)	67.3 ± 13.09 5/45	72.39±14.13 3/47	74.2 ± 13.85 0/50	>0.05 <sup>a</sup> >0.05 <sup>b</sup>
Weight	77.8± 21.3	75.46 ± 28.22	72.15 ± 18.77	>0.05 <sup>a</sup>
Height	168.1±10.1	169.56 ±10.02	170.6 ±11.5	>0.05 <sup>a</sup>
Heart rate (stand)	71.98±15.011	64.7 ± 12.81	61.56 ± 5.88	>0.05 <sup>a</sup>
Heart rate (sit)	80.14 ± 16.91	74 ± 15.504	68.48 ± 5.4	>0.05 <sup>a</sup>

Table 5.1 (Continued)

UPDRS-1	$28.94 \pm 15.09$	$16.14 \pm 3.07$	-	$<0.05^a$
UPDRS-2	$28.7 \pm 15.92$	$18 \pm 4.29$	-	$<0.05^a$
UPDRS-3	$65.64 \pm 36.37$	$31.4 \pm 7.82$	-	$<0.05^a$
UPDRS-4	$14.6 \pm 6.16$	$8.28 \pm 2.29$	-	$<0.05^a$
Total UPDRS score	$118.28 \pm 40.62$	$65.54 \pm 11.10$	-	$<0.05^a$
MoCA score	$15.04 \pm 2.18$	$21.28 \pm 2.20$	$28.28 \pm 1.249$	$<0.05^a$
<b>DaTscan (SBR features)</b>				
Right caudate				
Left caudate	$1.1 \pm 0.39$	$1.61 \pm 0.50$	$2.10 \pm 0.83$	$<0.05^a$
Right putamen	$1.26 \pm 0.48$	$1.56 \pm 0.47$	$2.00 \pm 0.97$	$<0.05^a$
Left putamen	$0.50 \pm 0.21$	$0.74 \pm 0.31$	$0.92 \pm 0.79$	$>0.05^a$
	$0.55 \pm 0.24$	$0.71 \pm 0.30$	$0.91 \pm 0.74$	$>0.05^a$
<b>Protein Biomarkers</b>				
$\alpha$ -synuclein (pg/ml)	$1927.56 \pm 51.47$	$1989.79 \pm 48.8$	$1874.76 \pm 42.7$	$<0.05^a$
$A\beta_{1-42}$ (pg/ml)	$884.5 \pm 313.84$	$1070.7 \pm 421.37$	$964.9 \pm 422.6$	$<0.05^a$
t-tau (pg/ml)	$17.6 \pm 66.25$	$19.80 \pm 66.08$	$18.7 \pm 71.96$	$<0.05^a$
P-tau <sub>181</sub> (pg/ml)	$14.36 \pm 6.31$	$15.89 \pm 5.02$	$17.27 \pm 8.18$	$<0.05^a$

Note: <sup>a</sup> = ANNOVA test, <sup>b</sup> = chi-square test

(UPDRS: Unified Parkinson's Disease Rating Scale, MoCA: Montreal Cognitive Assessment, SBR: Specific Binding Ratio, DaTscan: Dopamine Transporter Scan)

#### 5.4 Proposed Methodology

The proposed diagnostic system integrates multimodal data using a three-tier architecture comprising: (1) a 1d-CNN for classification, (2) a fine-tuned LLM for interpretability and personalised interaction, and (3) a cloud-based deployment interface for clinical integration. This section outlines the design principles, data processing pipeline, and model configurations adopted to build an end-to-end, scalable, and explainable framework for PD diagnosis and

monitoring. Multimodal inputs including MRI-based radiomics, DaT-derived SBR values, CSF biomarkers, and clinical scores are first pre-processed and normalized to create a unified feature vector. Ratio-based engineered features are also introduced to enhance biological interpretability and early-stage classification. Class imbalance, particularly due to limited prodromal cases, is addressed through synthetic oversampling techniques applied to the training set.

The processed features are passed into a custom-built 1d-CNN that performs multiclass classification (control, prodromal, PD). Outputs from the CNN, including interpretability scores generated via XAI techniques, are then utilised to fine-tune a domain-specific LLM based on ChatGPT-4.0. This LLM acts as a contextual interface, enabling natural language interaction, diagnostic summary generation, and clinical guidance. Both components are integrated into a secure, cloud-based platform that handles real-time inference, user interaction, and report management. This architecture aims to bridge the gap between high-performance AI algorithms and real-world usability in clinical settings, offering both predictive accuracy and interpretability in a deployable format.

#### **5.4.1 Proposed 1D-CNN model**

The 1d-CNN architecture developed in this study is optimized for analyzing structured, sequential data derived from multimodal sources. Unlike 2d-CNNs commonly used in image analysis, the 1D-CNN is particularly suited for feature vectors that encode biomarker levels, radiomic features, and clinical scores over a single dimension. The proposed network accepts a unified feature vector that includes: Radiomics features from MRI, SBR values from DaT-SPECT scans, Quantitative CSF protein biomarker levels, and UPDRS and MoCA clinical scores. The architectural design, as shown in Figure 5.1, begins with an input layer feeding into three consecutive 1D convolutional layers, with filter sizes of 32, 64, and 128 respectively. Each convolution layer employs a kernel size of 3 and uses the Leaky ReLU activation function defined as  $f(x) = \max(0.01x, x)$  to introduce non-linearity while avoiding vanishing

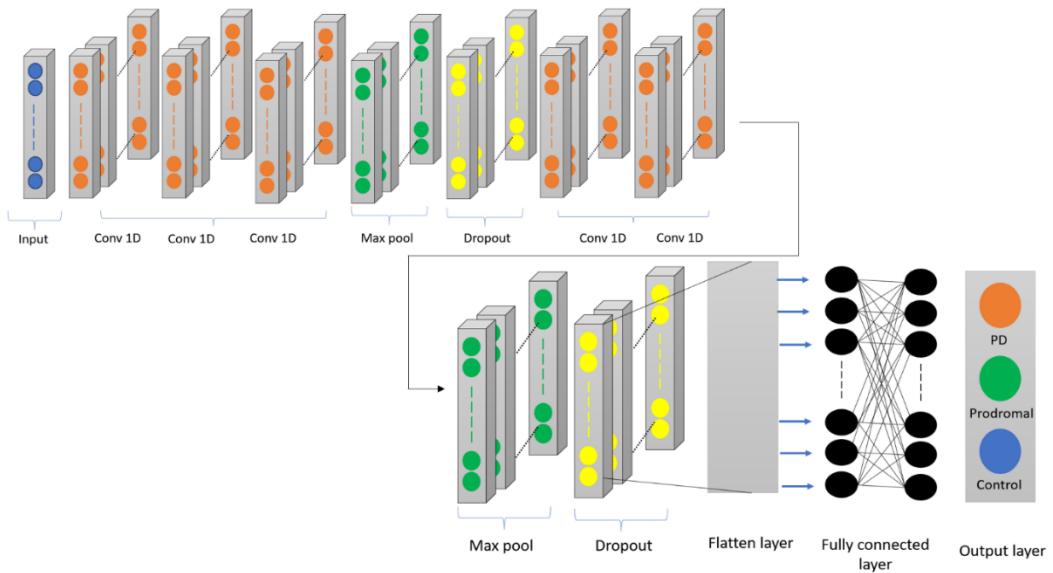
gradients. Mathematically, the 1D convolution operation for the  $i$ -th filter is defined as in Eq.5.1:

$$y^{(i)}[t] = \sum_{k=1}^K w_k^{(i)} \cdot x[t+k-1] + b^{(i)} \quad (5.1)$$

where  $y^{(i)}[t]$  represents the output at position  $t$ ,  $w_k^{(i)}$  are the filter weights,  $x[t]$  is the input signal, and  $b^{(i)}$  is the bias term. Each convolution block is followed by a max pooling layer (pool size = 2) to reduce dimensionality and preserve dominant features. Dropout layers with a rate of 0.5 are applied after pooling layers to mitigate overfitting. Two additional convolutional layers further refine the feature representation, after which a flattening operation is used to transform the output into a fully connected layer. The final output layer is a Softmax classifier designed for three-class classification (PD, prodromal, and control). The network is trained using categorical cross-entropy loss, computed as in Eq.5.2:

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\widehat{y_{ij}}) \quad (5.2)$$

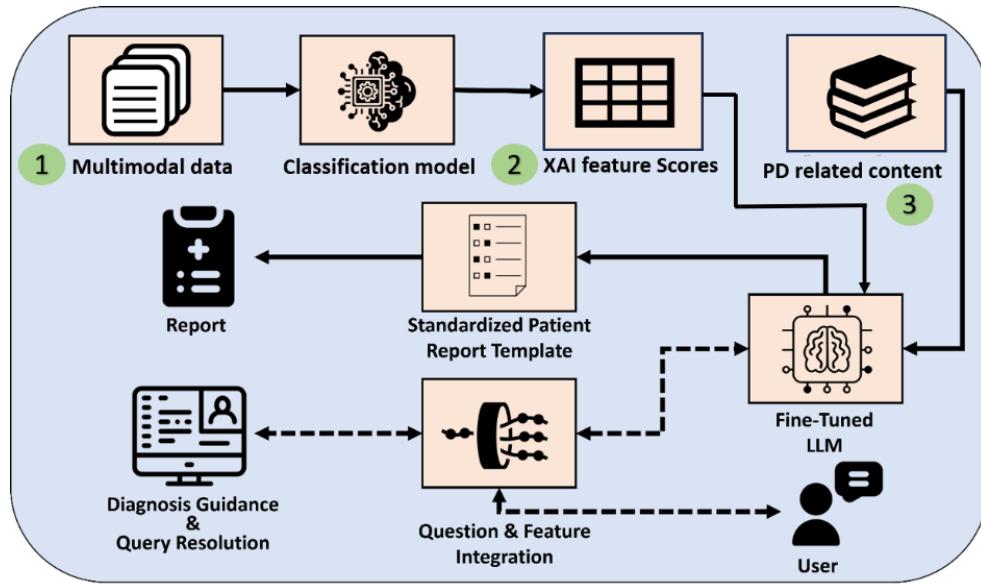
where  $N$  is the number of samples,  $C$  is the number of classes,  $y_{ij}$  is the true label,  $\widehat{y_{ij}}$  is the predicted probability. This architecture is capable of learning complex interdependencies between features across different modalities, offering better performance for early and differential PD classification.



**Figure 5.1.** Detailed architecture of the classification (1D-CNN) layer

#### 5.4.2 Fine-tuned Large language model for PD analysis

While CNN-based models excel at pattern recognition, they cannot often explain decisions or interact meaningfully with end users. To bridge this gap, a fine-tuned LLM based on ChatGPT-4.0 is integrated into the framework, enabling natural language interpretation of diagnostic outputs and real-time clinical interaction [119], [120]. The LLM is trained using PD-specific prompt-completion pairs derived from literature, medical forms, and clinical protocols. It also ingests structured inputs such as XAI-derived feature scores and classification outputs, allowing it to generate context-aware responses. The fine-tuning process incorporates RLHF to ensure safety, accuracy, and alignment with medical communication standards [31]. As shown in Figure 5.2, the LLM workflow comprises: (1) input embedding of multimodal features, (2) transformer-based reasoning for generating reports and answers, and (3) an interactive interface for user engagement. The model supports multi-turn conversations, handles PD-related queries, and rejects irrelevant inputs to maintain clinical focus. By transforming model outputs into explainable narratives and supporting dynamic dialogue, the LLM enhances interpretability, fosters trust, and adds a human-in-the-loop dimension to AI-assisted PD diagnosis.



**Figure 5.2** Finetuned LLM model for diagnosis of PD

### 5.4.3 Cloud-based interactive Health inquiry system

To translate the proposed diagnostic framework into a scalable and accessible clinical tool, this study incorporates a cloud-based interactive health inquiry system that serves as the deployment interface for both the 1d-CNN classifier and the fine-tuned LLM. This system is designed to bridge the gap between advanced AI algorithms and real-world clinical usability, offering seamless data interaction, real-time inference, and patient-specific medical insights. Cloud-based platforms have emerged as a powerful enabler in digital healthcare by offering scalability, data security, and remote accessibility [27]. In the context of PD diagnosis, cloud deployment ensures that patients and clinicians can engage with the system regardless of geographic or infrastructural constraints. This is particularly beneficial in under-resourced or rural areas where access to neurologists and imaging experts is limited [121].

The proposed system architecture enables users, patients, caregivers, or medical professionals to securely upload T2-weighted MRI scans, SPECT data, and clinical test results (e.g., UPDRS

and MoCA scores) through a user-friendly web interface. The backend cloud infrastructure handles all computational workloads, including image preprocessing, multimodal feature extraction, 1D-CNN classification, and LLM-based interpretation. Once data is processed, the system automatically generates: (1) Diagnostic predictions (PD, prodromal, or control), (2) Probabilistic confidence scores, (3) Biomarker analysis reports, (4) Natural language summaries of diagnostic outcomes, and (5) Personalized suggestions for clinical follow-up or lifestyle modifications. The cloud interface also supports real-time query resolution, where users can engage with the LLM to ask medically relevant questions. For example, a patient may inquire about the implications of their SBR values, while a clinician may request an explanation for the model's confidence in a prodromal classification. The LLM responds using structured medical knowledge and contextual information derived from the patient's uploaded data. Additionally, the platform implements end-to-end data encryption and role-based access control, ensuring compliance with privacy standards and clinical data governance protocols. The system is also designed to continuously improve diagnostic accuracy through feedback-driven model updates based on anonymized case histories and user interactions. By integrating advanced AI models within a robust cloud infrastructure, the proposed health inquiry system not only enhances diagnostic precision but also democratizes access to expert-level insights. This interactive, real-time capability transforms the traditional diagnostic pipeline into a responsive and patient-centred clinical tool, paving the way for practical deployment in both primary care and specialized neurological settings.

## 5.5 Results and Analysis

This section presents the experimental results obtained from the proposed multimodal diagnostic framework, which integrates a 1d-CNN classifier with a fine-tuned LLM. The analysis is divided into two key phases: (1) statistical evaluation of multimodal feature distributions, and (2) performance assessment of the classification model using various combinations of input features. The outcomes demonstrate the effectiveness of the system in detecting and distinguishing between PD, prodromal, and control cases with high precision.

### 5.5.1 Statistical Analysis of Multimodal Data

A comprehensive statistical analysis was conducted to evaluate the significance and discriminative power of each modality-specific feature across the three subject groups. Descriptive statistics (mean, standard deviation) were calculated for all numerical variables, while group differences were assessed using analysis of variance (ANOVA) for continuous features and chi-square tests for categorical variables such as gender and PD history. A threshold p-value of 0.05 was used to determine statistical significance. Out of the 21 initial features, 14 were found to be statistically significant and selected for further processing. These included key clinical metrics (UPDRS and MoCA scores), DaT-SPECT derived specific binding ratios (SBR), and protein biomarkers such as  $\alpha$ -synuclein,  $A\beta_{1-42}$ , t-Tau, and p-Tau<sub>181</sub>. Features such as height, weight, and heart rate, which did not meet the statistical threshold ( $p > 0.05$ ), were excluded from the subsequent modelling pipeline.

Table 5.1 provides a detailed summary of the multimodal features, including their distributions across the three groups and associated p-values. The analysis highlights that  $\alpha$ -synuclein and  $A\beta_{1-42}$  exhibit the highest intergroup variation, followed by UPDRS-3 and SBR values in the caudate regions, supporting their relevance in PD characterisation. This selection of statistically robust features ensures that the input to the classification model is both clinically meaningful and computationally efficient, thereby improving the overall reliability and interpretability of the system.

### 5.5.2 Performance evaluation of 1D-CNN classifier

The 1d-CNN classifier was trained and tested on a dataset composed of 150 subjects using a 70:30 split for training and validation. A total of 121 features were input to the model, comprising 14 statistically validated clinical, SPECT, and CSF features, and 107 radiomic features extracted from MRI scans. Before model training, min–max normalisation was applied to scale each feature within the range of 0 to 1, mitigating the effects of scale variance across modalities. The normalization is defined as Eq. 5.3.

$$X'_i = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (5.3)$$

Additionally, feature engineering was used to generate five biologically relevant ratio-based features, including: P-tau<sub>181</sub> / Total-tau, Total-tau / Aβ<sub>1-42</sub>, P-tau<sub>181</sub> / Aβ<sub>1-42</sub>, Right caudate / Left caudate and Right putamen / Left putamen. These ratios were designed to capture relative variations and hemispheric asymmetries associated with disease progression, especially in prodromal and early-stage PD cases [10], [122]. To address class imbalance, data augmentation was applied to the minority class (prodromal group) in the training set using synthetic oversampling. The performance of the classifier was evaluated using standard metrics accuracy, precision, recall (sensitivity), and F1-score calculated across different combinations of input features. The mathematical formulations are shown in Eq. 5.4 to Eq. 5.7. Here, TP is true positive, TN is true negative, FP is false positive, FN is false negative.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (5.4)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (5.5)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (5.6)$$

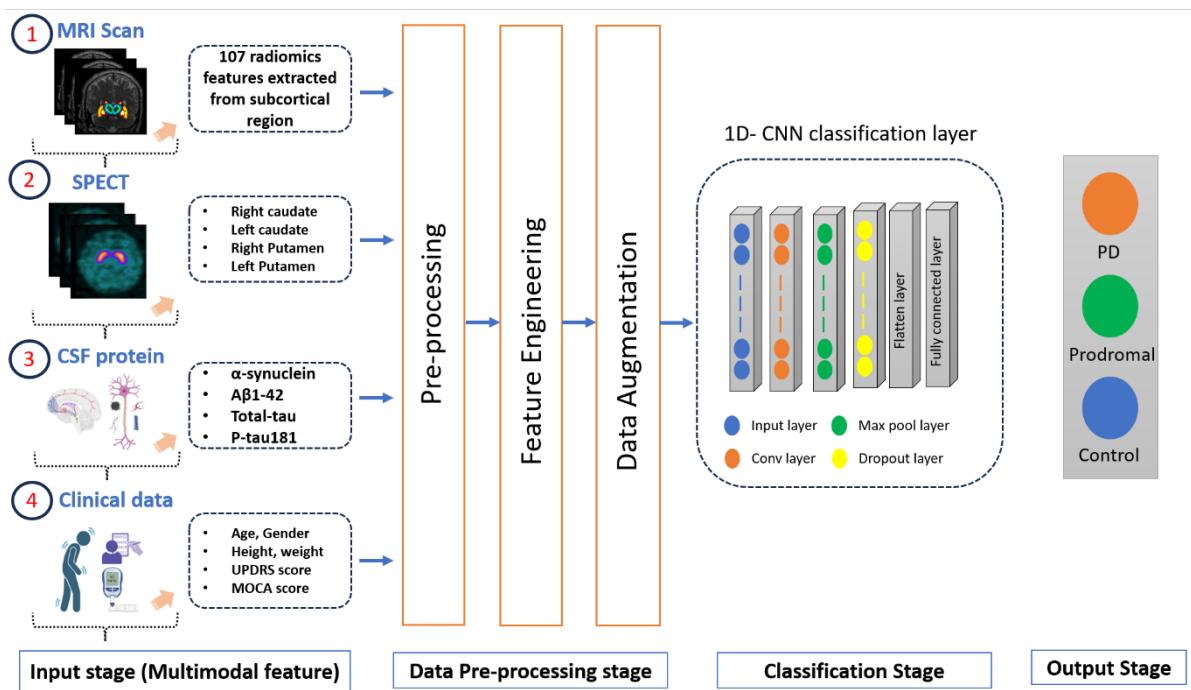
$$F1\ score = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \quad (5.7)$$

Table 5.2 presents the performance results using different feature combinations. The inclusion of MRI features improved accuracy to 92.7%, and the addition of engineered ratio features further enhanced the classification accuracy to 96.9%, with corresponding improvements in recall, precision, and F1-score across all classes. This evaluation confirms that the fusion of radiological, clinical, and biochemical modalities, complemented by derived features, substantially enhances diagnostic accuracy. The proposed 1d-CNN model exhibits strong generalization capability and clinical relevance, validated through multi-metric benchmarking

and visualization provided in Figure 5.3.

### 5.5.3 LLM-based diagnosis guidance and query resolution

To complement the diagnostic capability of the 1d-CNN classifier and enhance clinical interpretability, a fine-tuned LLM adapted from Chatgpt-4.0 was integrated into the system to facilitate dynamic, context-aware diagnosis guidance and query resolution. This model is designed not only to explain the predictions made by the DL model but also to provide patients and clinicians with medically grounded recommendations, tailored interpretations, and actionable insights.



**Figure 5.3.** The overview workflow of the 1d- CNN classifier model

**Table 5.2:** Comparison of features for multiclass classification in PD

S. No	Features	Accuracy (in %)	Recall (in %)	Precision (in %)	F1-score (in %)
1.	SPECT + CSF proteins + Clinical data	90.4	91.7	89.3	92.3
2.	MRI data	92.7	90.8	93.7	94.9
3.	SPECT + CSF proteins + Clinical data + MRI data	94.9	93.6	95.2	96.4
4.	All + Ratio based features	96.9	94.4	97.2	96.5

The fine-tuning process involved curating a domain-specific dataset that captured diverse clinical scenarios and diagnostic patterns relevant to PD. This included structured prompt-completion pairs generated from validated PD medical literature, diagnostic reports, and query-answer sets sourced from the “Expanded Library for Parkinson’s Disease Prompts”. To ensure clinical fidelity, the training process incorporated RLHF, along with real-world patient queries and XAI outputs derived from the CNN mode.

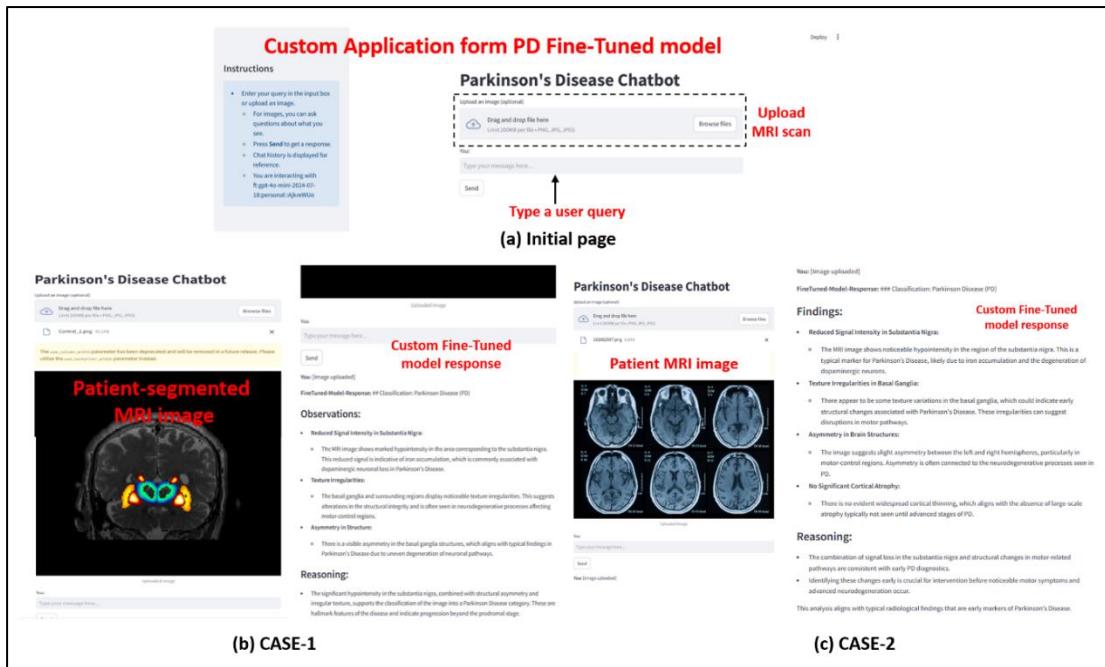
The dataset was structured using ScaleXI, a semi-automated annotation tool that efficiently extracted and formatted multimodal medical content. Each record in the training dataset mapped clinical features such as SBR values, MRI-based radiomic descriptors, CSF protein levels, and UPDRS/MoCA scores to diagnostic conclusions and medically relevant textual explanations. The LLM was optimized to accept these features in natural language prompts and return medically sound, user-understandable responses. As illustrated in Figure 5.4, the fine-tuned LLM is capable of processing image-derived inputs and structured clinical data to produce tailored text-based responses. These include summaries of abnormal brain patterns, interpretations of dopaminergic deficiencies (e.g., asymmetries in SBR values), and correlations between biomarker levels and disease severity. For instance, if a user uploads an

MRI scan and asks whether the image suggests prodromal PD, the model can interpret extracted features and explain which aspects (e.g., subtle subcortical shrinkage or asymmetric binding in DaTscan) are indicative of early-stage PD. The model also supports interactive, multi-turn query resolution, enabling users to follow up with clarifying questions. This dynamic exchange emulates a clinical dialogue, empowering both patients and medical professionals to explore the reasoning behind diagnostic predictions. Moreover, the model is designed to maintain focus on PD-specific content; off-topic questions are either respectfully declined or redirected to appropriate domains, preserving the integrity and scope of interaction.

As shown in Figure 5.5, the system handles both visual and textual queries effectively, leveraging integrated multimodal reasoning to respond with clinically relevant insights. This capability is critical for real-world deployment, where users may present a combination of structured records, imaging data, and natural language queries. In summary, the fine-tuned LLM enhances the transparency, interpretability, and accessibility of the AI framework. By combining deep neural inference with conversational intelligence, it transforms the diagnostic pipeline into an interactive, explainable, and patient-centric platform, supporting decision-making in both primary and specialized care settings.

#### **5.5.4 LLM-Driven personalised medical report generation**

A key feature of the proposed framework is its ability to generate personalized medical reports using the fine-tuned model. These reports integrate clinical data, radiological insights, and biochemical indicators into a structured and easily interpretable summary tailored to each patient. By leveraging the interpretive capabilities of the LLM, the system ensures that diagnostic outcomes are communicated in a meaningful, clinically grounded, and patient-friendly format.



**Figure 5.4.** Image-based example cases and corresponding Fine-tuned LLM responses

**User query 1**

You: Hello! Can you describe how changes in radiomic features in subcortical region from MRI scans contribute to early diagnosis of Parkinson's disease? **User query 1**

FineTuned-Model-Response: ## Changes in Radiomic Features from MRI Scans

**Contribution to Early Diagnosis of Parkinsons Disease**

- Subcortical Region MRI Analysis: **Custom Fine-Tuned model response**
  - Radiomic features in the subcortical regions, such as the putamen and caudate nucleus, are critical for detecting early signs of dopaminergic dysfunction.
  - These features help identify subtle changes not visible to the naked eye, allowing for earlier diagnosis and intervention.
- Quantitative Assessment:
  - MRI scans can quantify the volume, shape, and texture of subcortical structures, providing insights into potential neurodegeneration.
  - Changes in these metrics, especially a reduction in putamen DATSCAN scores below 0.8, indicate significant dopaminergic dysfunction.
- Detection of Prodromal Signs:
  - Radiomics can identify early stages where patients have mild cognitive changes MOCA ≥ 26 and subtle motor symptoms not yet severe enough to meet clinical UPDRS criteria.
  - This helps classify patients into the prodromal group, allowing for closer monitoring and potential early intervention.

**Case Scenario**

- Example: A patient with a UPDRS score of 1 and MOCA score of 27, but MRI shows changes consistent with mild dopaminergic dysfunction in the putamen, could be identified as being in the prodromal group through radiomic analysis.

**References**

- Stewart A. Factor, "Parkinsons Disease", for detailed information on classification criteria and early diagnosis approaches.
- Publication on radiomics in neurodegenerative diseases for additional insights beyond clinical evaluation.

**User query 2**

You: How do the interactions between radiomics and DATSCAN imaging improve diagnosis in Parkinson's patients? **User query 2**

FineTuned-Model-Response: Interactions Between Radiomics and DATSCAN Imaging in Parkinsons Disease Diagnosis **model response**

- Radiomics and DATSCAN Imaging
  - Radiomics* involves extracting a large number of features from medical images using data-characterization algorithms. It helps in identifying subtle patterns not visible to the naked eye.
  - DATSCAN* is a specialized imaging technique using SPECT to visualize dopamine transporter activity in the brain, crucial for diagnosing Parkinsons Disease.
- Improvement in Diagnosis
  - Complementary Information:* DATSCAN provides direct evidence of dopaminergic dysfunction, while radiomics can reveal structural and morphological changes in the brain.
  - Enhanced Sensitivity and Specificity:* Combining radiomics with DATSCAN improves the ability to detect early-stage Parkinsons, potentially even before clinical symptoms manifest.
  - Assessment of Disease Progression:* Radiomics can help monitor changes over time, providing insights into disease progression and response to therapy. DATSCAN helps determine the severity of motor-related complications.
- Example or Case Scenario
  - A patient shows mild cognitive decline and subtle motor symptoms, but standard clinical evaluation is inconclusive. DATSCAN reveals significant dopaminergic dysfunction, and radiomic analysis identifies structural changes in basal ganglia. This combination leads to an earlier and more accurate diagnosis of Parkinsons Disease.

**References**

- Stewart A. Factor. Parkinsons Disease Diagnosis and Management.

**User query 3**

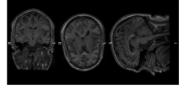
You: 6. How does the progression of amyotrophic lateral sclerosis (ALS) differ across genetic and sporadic cases? **User query 3**

FineTuned-Model-Response: "Your query is not related to Parkinson's Disease." **model response**

**Figure 5.5.** Text-based examples cases and their corresponding finetuned model responses

Once the multimodal data comprising MRI scans, SPECT-derived SBR values, CSF biomarker levels, and clinical assessment scores is processed through the diagnostic pipeline, the LLM synthesizes the results into a comprehensive report. This includes: (1) A brief overview of the patient's demographic and clinical profile, (2) Interpretation of neuroimaging findings (e.g., volumetric shrinkage, asymmetries), (3) Biochemical marker evaluation (e.g., abnormal protein levels), (4) An assessment of motor and cognitive scores (e.g., UPDRS, MoCA), and (5) Final diagnostic impressions and follow-up recommendations.

As depicted in Figure 5.6, the platform interface captures patient information through structured input fields. Following image segmentation and feature extraction, the LLM dynamically generates contextual descriptions of the uploaded MRI scans. It cross-references this data with clinical scores and biomarker trends to generate condition-specific interpretations such as identifying features consistent with prodromal PD, confirming a mild or advanced PD diagnosis, or ruling out PD-related anomalies in control subjects. Furthermore, the report includes suggested next steps for both clinicians and patients. These may consist of additional lab tests, lifestyle adjustments, or therapeutic strategies, depending on the model's interpretation of the disease stage. For example, a subject flagged as prodromal may be advised to undergo frequent neuropsychological monitoring and dopamine reuptake imaging, while a confirmed PD case may receive recommendations on medication titration and physiotherapy. Each report is generated in real-time and is securely stored within the cloud interface, enabling longitudinal tracking and comparative analysis during future visits. The integration of LLM-based narrative generation ensures that the technical results of the CNN model and imaging analysis are distilled into clinically actionable summaries, thereby enhancing the accessibility of AI-driven diagnosis across user groups. In essence, this component bridges the gap between algorithmic output and real-world usability, transforming raw diagnostic data into structured reports that are informative for clinicians and comprehensible to patients. This facilitates collaborative decision-making and supports continuous engagement in patient care pathways.

MEDICAL REPORT			
<b>PERSONAL DATA</b>			
NAME	SIRIYA S	PATIENT ID.	216
AGE	47	SEX	FEMALE
VISIT NO.	08	BLOOD GROUP	A1 +VE
HEIGHT	172 CMS	WEIGHT	65 KGS
<b>MEDICAL HISTORY</b>			
MEDICATION(S)	LEVODOPA, MAO-B INHIBITORS		
SURGERIES	DBS		
HISTORY OF PD	YES		
<b>DIAGNOSIS RESULTS</b>			
①		②	
① MRI INPUT	② Segmented Subcortical Region		
IMAGE DESCRIPTION(S)	THICKNESS CHANGES IN SN REGIONS, CAUDATE & PUTAMEN IS LESS SHAPED, INFLAMMATION IN THALAMUS		
SUGGESTED LAB TEST(S)	BLOOD TEST DATSCAN		
<b>PREDICTIVE ANALYSIS</b>			
CSF - APLHA - SYNUCLEIN	2050.9 PG/ML	PATIENT ID.	216
T-TAU	29.45	SEX	FEMALE
P-TAU	97.05	UPDRS - III	60
MOCA	32	UPDRS - II	2

PAGE 1/2

Disclaimer: This report generated by a computerized AI-based platform is in its beta version and may undergo further refinement and validation.

**Figure 5.6** Report generated by the fine-tuned LLM using a standardized diagnostic template

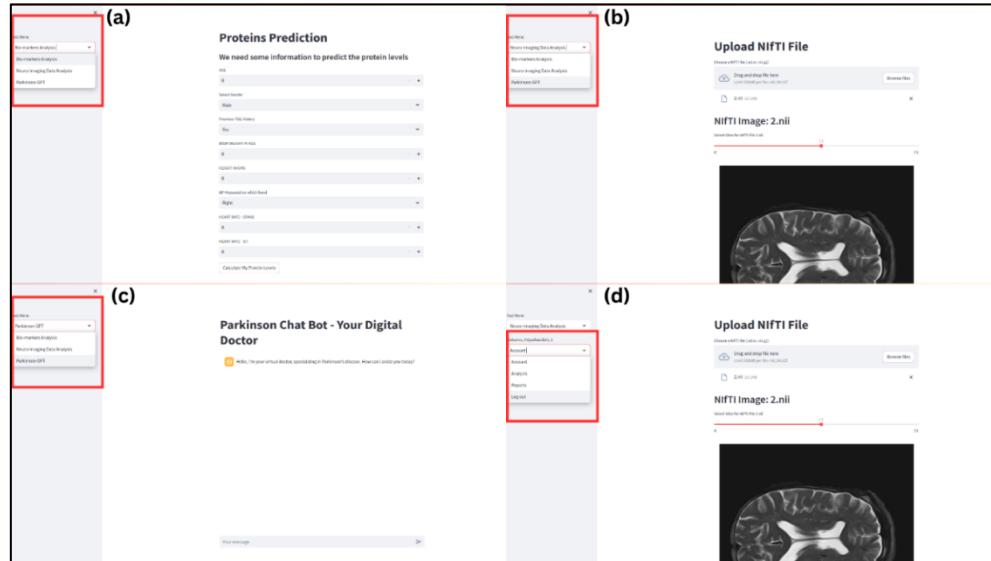
### 5.5.5 Integrated Cloud-Based Comprehensive Record Management

The cloud-based record management system embedded within the proposed diagnostic framework plays a pivotal role in unifying all stages of patient evaluation data entry, image analysis, prediction, and personalized reporting within a secure, interactive, and user-friendly interface. This system not only centralizes multimodal diagnostic data but also supports real-time interaction and longitudinal record tracking, making it suitable for routine clinical deployment. The design of the cloud platform prioritizes usability, modularity, and interpretability, offering distinct functional modules accessible via a consistent navigation layout. As illustrated in Figure 5.7, each module is crafted to fulfill a specific aspect of the diagnostic pipeline:

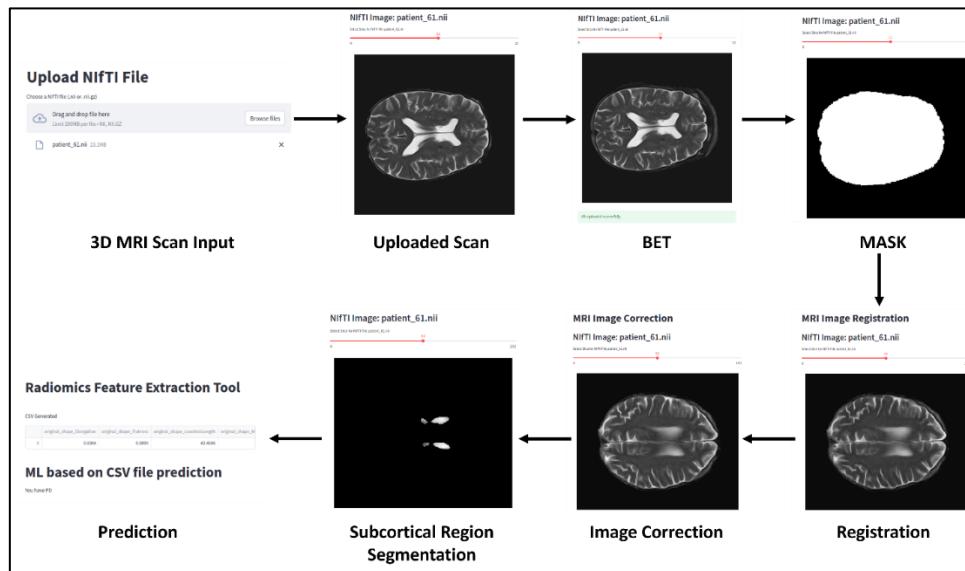
Figure 5.7(a) shows the Proteins Prediction Module, where users can input relevant clinical parameters such as age, gender, PD history, UPDRS scores, and heart rate. Based on this structured input, the platform estimates CSF protein levels—including  $\alpha$ -synuclein, A $\beta$ 1-42, tTau, and pTau181 via machine learning models trained on clinical-biochemical correlations. Figure 5.7(b) presents the Brain Imaging Data Analysis Module, which allows users to upload neuroimaging data in NiFTI format. Once uploaded, these images are subjected to preprocessing steps such as registration, segmentation, and radiomics-based feature extraction, forming the input for the CNN-based diagnostic prediction. Figure 5.7(c) illustrates the Parkinson Chat Bot Module, powered by a fine-tuned LLM. This AI-driven digital assistant supports natural language conversations with users, providing clarification, interpretation of results, and clinical suggestions related to PD. Figure 5.7(d) displays the User Navigation Menu, offering quick access to key platform functionalities including account management, imaging analysis, result generation, and AI report download, thereby supporting a seamless end-user experience.

Further extending the functionality of the platform, Figure 5.8 provides a comprehensive view of the end-to-end image processing and inference workflow. It demonstrates the step-by-step progression from uploading a Nifti image, performing brain segmentation and registration, through to radiomics feature extraction and ML-based classification. This automated pipeline ensures consistency in preprocessing and interpretability of outputs, ultimately leading to a prediction result indicating the presence or absence of PD. In this example, the platform not only handles image correction and registration tasks but also outputs a CSV file of extracted features. These features feed into a pre-trained classification model, which, in this case, outputs a diagnostic conclusion "You have PD" based on the analysed data. This complete flow, visualised in Figure 5.8, reinforces the system's ability to provide accurate and reproducible diagnoses through transparent steps visible to clinicians and researchers. Collectively, Figures 5.7 and 5.8 showcase the modular, explainable, and patient-centric design of the cloud-based system. By providing integrated access to imaging tools, clinical analytics, LLM-powered

interaction, and automated reporting, the platform supports continuous monitoring and decision-making, empowering clinicians to track disease progression while ensuring patients receive timely, accessible insights into their neurological health.



**Figure 5.7** User interface of the proposed cloud-based platform (a) Proteins prediction module interface, allowing users to input clinical data for predicting protein levels related to PD. (b) Brain imaging data analysis interface, supporting the upload and processing of NiFTI files for advanced analysis. (c) Parkinson Chat Bot interface, acting as a digital assistant specialized in PD. (d) User navigation dropdown showcasing available tools within the platform to access various functionalities.



**Figure 5.8** Brain imaging data analysis interface, supporting the upload and processing of NiFTI files for advanced analysis

## 5.6 Discussion

### 5.6.1 Datatypes for finetuning and Testing

A total of 1,000 prompt-completion pairs were curated to fine-tune the LLM. This dataset comprised 250 manually collected entries, 250 pairs extracted from the “Expanded Library for Parkinson’s Disease Prompts,” and 500 Q&A pairs generated using the ScaleXI package. These structured prompt-completion datasets were derived from contextual inputs and designed to reflect real-world clinical relevance. Each prompt was carefully formulated based on predefined types including open-ended, binary (yes/no), demographic, and classification-based queries to ensure diversity and clinical comprehensiveness. Completions were generated via OpenAI’s API to maintain contextual accuracy and linguistic coherence. The final dataset adhered to a standardized JSON/CSV format, ensuring compatibility with fine-tuning and evaluation pipelines. Out of the 1,000 entries, 915 pairs were used for fine-tuning and 85 for testing, ensuring robust learning and generalizability. These entries covered a wide range of PD-relevant data domains, as detailed in Table 5.3, ensuring the model was exposed to a representative distribution of feature types and clinical scenarios.

**Table 5.3:** PD-relevant data domains Q/A pairs for fine-tuned and testing the model

Data Type	Q/A Pairs (Fine-tuned Model)	Q/A Pairs (Testing)
Clinical data	194	10
Protein data	174	18
MRI data	151	11
SPECT	196	9
XAI feature score	78	13
PD related content	122	24
Total	915	85

### **5.6.2 Parameter Fine-Tuning LLM**

The LLM fine-tuning process was focused on increasing the accuracy and reliability of responses to PD-related queries while minimizing hallucinations. The goal was to ensure that the model remains grounded in verified clinical knowledge and produces precise, domain-specific outputs. The fine-tuning was performed on GPT-4 (1 trillion parameters), guided by a custom system prompt that directed the model to operate as a concise and accurate PD-focused chatbot. To promote learning stability and efficiency, the training was conducted over three epochs, with careful optimisation of the hyperparameters. Key fine-tuning parameters were selected to strike a balance between computational efficiency and model performance. Each input and output sequence were limited to 256 tokens, for a total sequence length of 512 tokens. The token-averaged cross-entropy loss function was employed along with the Adam optimizer at a learning rate of 0.0005. A batch size of 16 was used for memory efficiency.

To ensure coherent and diverse outputs, generation-specific parameters such as temperature (0.3) and repetition penalty (1.2) were configured. The ScaleXI Python package was employed to automate data preprocessing, tokenization, cost estimation, and performance evaluation, significantly streamlining the end-to-end training process. Fine-tuning was executed via OpenAI's fine-tuning API, which handled the Dataset uploads, Training job initiation, Progress monitoring, Deployment of the fine-tuned model for real-world usage. These optimizations led to a highly specialized and clinically reliable LLM capable of generating expert-level responses to PD-related patient and clinician queries.

### **5.6.3 Evaluation method for finetuned LLM response**

Following the fine-tuning of the GPT-4o-mini model, it was essential to rigorously assess its output quality. The evaluation aimed to ensure that the model consistently delivered accurate, medically relevant, and interpretable responses. A three-stage automated evaluation pipeline was implemented using GPT-4 as the LLM-based evaluator.

### 5.6.3.1 Structured Prompt-Completion pairs for evaluation

The evaluation dataset was stored in JSONL format, with each record comprising a medical prompt representing the user query, a ground truth completion verified by domain experts or reliable PD knowledge bases, and the model-generated completion from the fine-tuned LLM. An example format is provided in Table 5.4, showcasing how factual and semantic alignment was quantitatively assessed. The automated evaluation system compared model outputs with ground truth responses using scoring criteria applied by GPT-4.

**Table 5.4:** Sample JSON format

Key	Value
"prompt"	"What are the early symptoms of Parkinson's disease?"
"ground_truth_completion"	"The early symptoms of Parkinson's include tremors, bradykinesia, rigidity, and postural instability."
"model_completion"	"Symptoms of Parkinson's can include hand tremors, stiffness, and slow movement."

### 5.6.3.2 Evaluating using GPT-4.0 Mini as a Judge

To ensure objective, consistent assessment, GPT-4 was configured as a neutral medical evaluator, scoring responses on a scale of 0.0 to 5.0. The model was instructed using a structured evaluation prompt that asked it to assess factual accuracy, Coherence, Specificity, and Relevance. A concise justification ( $\leq 50$  words) accompanied each score, with the output formatted in CSV for structured analysis. Table 5.5 outlines the grading criteria used in scoring.

**Table 5.5:** Evaluation criteria for generated response

Score (0.0-0.5)	Evaluation Criteria
5.0	Perfect response - fully aligned with the ground truth.
4.0 - 4.9	Mostly correct - minor missing details
3.0 - 3.9	Partially correct – some factual errors or missing elements
2.0 - 2.9	Limited relevance- significant omissions or inaccuracies
1.0 - 1.9	Incorrect but related – contains relevant concepts but is factually incorrect.
0.0 - 0.9	Completely off-topic or misleading response.

### 5.6.3.3 Three-step Automated evaluation process

The evaluation framework followed a systematic three-step process: In Step 1, the fine-tuned GPT-4o-mini model was applied to the test dataset. Output generation was performed using the previously optimized generation parameters: temperature (0.3) and repetition penalty (1.2). In Step 2, generated completions were compared against ground truth completions to evaluate semantic and factual consistency. Finally, in Step 3, GPT-4 was employed as an independent evaluator, scoring each response and providing concise justifications. Each evaluation output was recorded in a CSV file containing user prompt, ground truth response, model-generated response, assigned score (0.0 – 5.0), and a brief justification comment ( $\leq 50$  words). This method allowed for a compact yet comprehensive quantitative and qualitative assessment of the LLM’s diagnostic communication capabilities.

## 5.7 Summary

- This chapter introduced an AI-powered diagnostic framework for Parkinson’s Disease that integrates multimodal data sources including neuroimaging, clinical assessments, and CSF biomarkers into a unified 1d-CNN classifier for accurate multiclass classification.

- A fine-tuned LLM was developed to enhance model explainability, offer real-time query resolution, and generate personalized diagnostic reports, making AI-driven predictions accessible and interpretable to both clinicians and patients.
- The cloud-based platform supported seamless integration of data upload, model inference, chatbot interaction, and longitudinal health record management, ensuring scalability, data security, and user-centric design.
- Experimental results demonstrated high classification accuracy (up to 96.9%) when combining radiomic features with clinical and biochemical data, highlighting the effectiveness of multimodal fusion and engineered ratio features in improving model performance.
- The LLM was rigorously fine-tuned and evaluated using a curated prompt-completion dataset and an automated GPT-4.0 mini-based judging mechanism, ensuring factual consistency, semantic alignment, and clinical relevance in generated responses.
- Overall, the proposed system represents a clinically viable, explainable, and scalable AI solution for early diagnosis, patient engagement, and continuous monitoring in PD care.

## CHAPTER 6 CONCLUSION AND FUTURE SCOPE

### **6.1 Conclusion**

Diagnosing Parkinson's Disease (PD) at an early stage and accurately distinguishing it from other neurodegenerative disorders remains a major challenge in clinical practice. This difficulty is primarily due to overlapping symptoms and varied clinical presentations, often resulting in misdiagnosis and delayed treatment. To address these limitations, this thesis proposes a comprehensive AI-driven diagnostic framework focused on early detection, personalised monitoring, and enhanced clinical interpretability through explainable models.

The proposed framework integrates radiomics, machine learning, deep neural networks, feature fusion strategies, and optimisation algorithms to address key challenges in early diagnosis, model transparency, and clinical applicability. The research is organised across three core chapters (Chapters 3, 4, and 5), each making substantial contributions to the fields of medical image analysis, explainable artificial intelligence (XAI), and multimodal data integration. Collectively, the methodologies and results presented in this thesis work establish a strong foundation for the development of intelligent, interpretable, and scalable diagnostic systems for PD.

- **Chapter 3** presented a machine learning-based diagnostic framework that utilises radiomics features extracted from 3d T2-weighted MRI scans, effectively addressing the limitations of conventional 2D analysis. A real-world dataset comprising 500 MRI scans from PD, prodromal, and healthy controls (sourced from the PPMI database) was employed. A comprehensive preprocessing pipeline includes brain extraction, registration, bias correction, normalisation, and segmentation. From the segmented subcortical regions, 107 radiomics features were initially extracted and subsequently reduced to the top 20 most predictive features using a two-stage feature selection method. Among several machine

learning models evaluated, the Gradient Boosting classifier achieved the highest performance, with an accuracy of 96.8%. To enhance clinical relevance, XAI techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model Agnostic Explanation (LIME) were employed, highlighting the diagnostic importance of Gray-Level Co-Occurrence Matrix (GLCM) texture features like Difference Entropy and Joint Entropy. These results provided not only high classification accuracy but also meaningful insights into early structural changes in PD.

- **Chapter 4** presents a deep learning-based method using a custom 3d Convolutional Neural Network (3d-CNN), which initially achieved 82.02% accuracy. After refining the architecture and tuning the parameters, the model reached 93.4% accuracy. A 3d-ResNet model was also tested and achieved 90% accuracy. Feature fusion using Canonical Correlation Analysis (CCA) combined the strengths of both models and raised the accuracy to 95%. Further optimization using the Whale Optimization Algorithm (WOA) enhanced the model performance to 97.2%, outperforming other metaheuristic approaches such as PSO, GA, GSO, and ACO. The proposed method demonstrated strong generalisability across different MRI scanner types (1.5T and 3T), confirming its robustness and potential for real-world clinical deployment.
- **Chapter 5** presents a multimodal AI framework that integrates clinical scores, imaging data, and biomarker information using a custom 1d-CNN model and a fine-tuned Chatgpt-4.0 interface for PD detection. The system combines 121 features, including SPECT-derived SBR values, CSF protein biomarkers, MRI-based radiomics, and clinical features into a unified prediction pipeline. The baseline 1d-CNN model achieved 94.9% accuracy, which was further improved to 96.9% with the inclusion of ratio-based biomarkers. To improve interpretability and user interaction, a domain-specific version of Chatgpt-4.0 was fine-tuned using structured PD-related prompts, explainable feature insights, and clinical knowledge. The language model supports multimodal input (text and image) and delivers personalised diagnostic reports, interactive responses, and detailed medical summaries. The

complete system was deployed on a secure, cloud-based platform with role-based access for clinicians, patients, developers, and administrators, ensuring scalability, usability, and accessibility in real-world healthcare environments.

Collectively, these chapters present a progressive and multidimensional approach to PD diagnosis. The developed AI frameworks not only demonstrate exceptional diagnostic accuracy but also significantly improve interpretability and clinical applicability. This work lays a solid foundation for future integration into personalised, AI-powered healthcare systems for neurodegenerative disease management.

## **6.2 Future scope**

While the proposed framework demonstrates significant potential and has achieved high diagnostic accuracy, robustness, and interpretability, several promising directions remain for further enhancement and clinical translation. In Chapter 3, future research should consider incorporating longitudinal neuroimaging data such as functional MRI (fMRI) and Diffusion Tensor Imaging (DTI) to improve the framework's ability to capture dynamic changes in brain function and structure. Additionally, expanding the dataset to include motor task-related assessments could help the model detect subtle motor impairments indicative of early PD progression.

In Chapter 4, further exploration of hybrid or ensemble deep neural network architectures may improve feature extraction and fusion performance, potentially leading to higher classification accuracy. Beyond the WOA, evaluating other bio-inspired or hybrid optimisation techniques could reveal more discriminative feature subsets and increase model efficiency. In Chapter 5, integrating direct feedback from expert neurologists into the training and refinement process of both the CNN and the language model would enhance clinical relevance and ensure that the system's outputs align closely with real-world diagnostic expectations. Future work should also explore real-time, continuous patient monitoring through wearable sensors and implement

diagnostic models on mobile or edge-computing platforms to enable proactive, remote tracking of motor and non-motor symptoms.

To support privacy-preserving collaboration across institutions, federated learning approaches could be employed. These would allow distributed model training without compromising sensitive patient data, facilitating scalable and secure adoption across healthcare centres. Ultimately, extensive clinical validation through user studies, multi-centre clinical trials, and long-term performance tracking will be essential to transition the developed diagnostic framework from research to routine clinical use. These efforts will help establish the framework as a reliable, interpretable, and accessible tool that significantly contributes to early PD diagnosis, personalised patient care, and improved healthcare.

## REFERENCES

- [1] G. Li *et al.*, “3D texture analyses within the substantia nigra of Parkinson’s disease patients on quantitative susceptibility maps and R2\* maps,” *Neuroimage*, vol. 188, no. October 2018, pp. 465–472, Mar. 2019, doi: 10.1016/j.neuroimage.2018.12.041.
- [2] K. Rosenberg-Katz, T. Herman, Y. Jacob, E. Kliper, N. Giladi, and J. M. Hausdorff, “Subcortical Volumes Differ in Parkinson’s Disease Motor Subtypes: New Insights into the Pathophysiology of Disparate Symptoms,” *Front. Hum. Neurosci.*, vol. 10, no. July, pp. 1–9, Jul. 2016, doi: 10.3389/fnhum.2016.00356.
- [3] N. Betrouni *et al.*, “Texture-based markers from structural imaging correlate with motor handicap in Parkinson’s disease,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, 2021, doi: 10.1038/s41598-021-81209-4.
- [4] J. M. Ranson, E. Kuźma, W. Hamilton, G. Muniz-Terrera, K. M. Langa, and D. J. Llewellyn, “Predictors of dementia misclassification when using brief cognitive assessments,” *Neurol. Clin. Pract.*, vol. 9, no. 2, pp. 109–117, 2019, doi: 10.1212/CPJ.0000000000000566.
- [5] S. Chaudhary *et al.*, “Domain specific cognitive impairment in Parkinson’s patients with mild cognitive impairment,” *J. Clin. Neurosci.*, vol. 75, pp. 99–105, May 2020, doi: 10.1016/j.jocn.2020.03.015.
- [6] L. Ma *et al.*, “Parkinson’s disease: Alterations in iron and redox biology as a key to unlock therapeutic strategies,” *Redox Biol.*, vol. 41, 2021, doi: 10.1016/j.redox.2021.101896.
- [7] W. Zeng, J. Cai, L. Zhang, and Q. Peng, “Iron Deposition in Parkinson’s Disease: A Mini-Review,” *Cell. Mol. Neurobiol.*, vol. 44, no. 1, 2024, doi: 10.1007/s10571-024-01459-4.
- [8] L. I. Wallis *et al.*, “MRI assessment of basal ganglia iron deposition in Parkinson’s disease,” *J. Magn. Reson. Imaging*, vol. 28, no. 5, pp. 1061–1067, 2008, doi: 10.1002/jmri.21563.

- [9] J. Parkinson, “An essay on the Shaking Palsy,” *Arch. Neurol.*, vol. 20, no. 4, pp. 441–445, 1969, doi: 10.1001/archneur.1969.00480100117017.
- [10] S. Hall, Y. Surova, A. Öhrfelt, H. Zetterberg, D. Lindqvist, and O. Hansson, “CSF biomarkers and clinical progression of Parkinson disease,” *Neurology*, vol. 84, no. 1, pp. 57–63, 2015, doi: 10.1212/WNL.0000000000001098.
- [11] Y. Xiao, S. Beriault, G. B. Pike, and D. L. Collins, “Multicontrast multiecho FLASH MRI for targeting the subthalamic nucleus,” *Magn. Reson. Imaging*, vol. 30, no. 5, pp. 627–640, Jun. 2012, doi: 10.1016/j.mri.2012.02.006.
- [12] H.-M. Zeng, H.-B. Han, Q.-F. Zhang, and H. Bai, “Application of modern neuroimaging technology in the diagnosis and study of Alzheimer’s disease,” *Neural Regen. Res.*, vol. 16, no. 1, p. 73, 2021, doi: 10.4103/1673-5374.286957.
- [13] D. Su *et al.*, “Projections for prevalence of Parkinson’s disease and its driving factors in 195 countries and territories to 2050: modelling study of Global Burden of Disease Study 2021,” *BMJ*, vol. 388, p. e080952, 2025, doi: 10.1136/bmj-2024-080952.
- [14] Y. Luo, L. Qiao, M. Li, X. Wen, W. Zhang, and X. Li, “Global, regional, national epidemiology and trends of Parkinson’s disease from 1990 to 2021: findings from the Global Burden of Disease Study 2021,” *Front. Aging Neurosci.*, vol. 16, no. January, p. 1498756, 2024, doi: 10.3389/fnagi.2024.1498756.
- [15] N. Tambasco *et al.*, “T2\*-weighted MRI values correlate with motor and cognitive dysfunction in Parkinson’s disease,” *Neurobiol. Aging*, vol. 80, pp. 91–98, Aug. 2019, doi: 10.1016/j.neurobiolaging.2019.04.005.
- [16] S. Marino *et al.*, “Magnetic resonance imaging markers for early diagnosis of Parkinson’s disease,” *Neural Regen. Res.*, vol. 7, no. 8, pp. 611–619, 2012, doi: 10.3969/j.issn.1673-5374.2012.08.009.
- [17] N. Aggarwal, B. S. Saini, and S. Gupta, “Feature engineering-based analysis of DaTSCAN-SPECT imaging-derived features in the detection of SWEDD and Parkinson’s disease,” *Comput. Electr. Eng.*, vol. 117, no. November 2023, p. 109241, 2024, doi: 10.1016/j.compeleceng.2024.109241.

- [18] L. Owolabi, “Progressive supranuclear palsy misdiagnosed as Parkinson’s disease: A case report and review of literature,” *Ann. Med. Health Sci. Res.*, vol. 3, no. 5, p. 44, 2013, doi: 10.4103/2141-9248.121221.
- [19] M. R. Salmanpour, M. Shamsaei, and A. Rahmim, “Feature selection and machine learning methods for optimal identification and prediction of subtypes in Parkinson’s disease,” *Comput. Methods Programs Biomed.*, vol. 206, p. 106131, Jul. 2021, doi: 10.1016/j.cmpb.2021.106131.
- [20] J. Zhang, “Mining imaging and clinical data with machine learning approaches for the diagnosis and early detection of Parkinson’s disease,” *npj Park. Dis.*, vol. 8, no. 1, 2022, doi: 10.1038/s41531-021-00266-8.
- [21] M. A. Islam, M. Z. Hasan Majumder, M. A. Hussein, K. M. Hossain, and M. S. Miah, “A review of machine learning and deep learning algorithms for Parkinson’s disease detection using handwriting and voice datasets,” *Heliyon*, vol. 10, no. 3, p. e25469, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25469.
- [22] M. Bakator and D. Radosav, “Deep Learning and Medical Diagnosis: A Review of Literature,” *Multimodal Technol. Interact.*, vol. 2, no. 3, p. 47, Aug. 2018, doi: 10.3390/mti2030047.
- [23] M. Obayya, M. K. Saeed, M. Maashi, S. S. Alotaibi, A. S. Salama, and M. A. Hamza, “A novel automated Parkinson’s disease identification approach using deep learning and EEG,” *PeerJ Comput. Sci.*, vol. 9, 2023, doi: 10.7717/peerj-cs.1663.
- [24] H. Zhai *et al.*, “Voxel-based morphometry of grey matter structures in Parkinson’s Disease with wearing-off,” *Brain Imaging Behav.*, vol. 17, no. 6, pp. 725–737, 2023, doi: 10.1007/s11682-023-00793-3.
- [25] M. T. Ribeiro, S. Singh, and C. Guestrin, ““Why Should I Trust You?,”” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.
- [26] S. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Section 2, pp. 4766–4775, May 2017, [Online]. Available: <http://arxiv.org/abs/1705.07874>

- [27] N. A. Hroub, A. N. Alsanna, M. Alowaifeer, M. Alfarraj, and E. Okafor, “Explainable deep learning diagnostic system for prediction of lung disease from medical images,” *Comput. Biol. Med.*, vol. 170, no. December 2023, p. 108012, 2024, doi: 10.1016/j.combiomed.2024.108012.
- [28] K. Asano and J. Chun, “Post-hoc explanation using a mimic rule for numerical data,” *ICAART 2021 - Proc. 13th Int. Conf. Agents Artif. Intell.*, vol. 2, no. Icaart, pp. 768–774, 2021, doi: 10.5220/0010238907680774.
- [29] M. T. Ribeiro and C. Guestrin, “Anchors : High-Precision Model-Agnostic Explanations”.
- [30] Y. Dai, Y. Gao, and F. Liu, “Transmed: Transformers advance multi-modal medical image classification,” *Diagnostics*, vol. 11, no. 8, pp. 1–15, 2021, doi: 10.3390/diagnostics11081384.
- [31] L. Ouyang *et al.*, “Training language models to follow instructions with human feedback,” *Adv. Neural Inf. Process. Syst.*, vol. 35, 2022.
- [32] OpenAI *et al.*, “GPT-4 Technical Report,” vol. 4, pp. 1–100, 2023, [Online]. Available: <http://arxiv.org/abs/2303.08774>
- [33] G. Solana-Lavalle and R. Rosas-Romero, “Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson’s disease,” *Comput. Methods Programs Biomed.*, vol. 198, 2021, doi: 10.1016/j.cmpb.2020.105793.
- [34] H. Zhai *et al.*, “Voxel-based morphometry of grey matter structures in Parkinson’s Disease with wearing-off,” *Brain Imaging Behav.*, no. 0123456789, 2023, doi: 10.1007/s11682-023-00793-3.
- [35] C. Yang, J. Chang, X. Liang, X. Bao, and R. Wang, “Gray Matter Alterations in Parkinson’s Disease With Rapid Eye Movement Sleep Behavior Disorder: A Meta-Analysis of Voxel-Based Morphometry Studies,” *Front. Aging Neurosci.*, vol. 12, no. August, pp. 1–10, 2020, doi: 10.3389/fnagi.2020.00213.
- [36] S. Chakraborty, S. Aich, and H.-C. Kim, “3D Textural, Morphological and Statistical Analysis of Voxel of Interests in 3T MRI Scans for the Detection of Parkinson’s Disease Using Artificial Neural Networks,” *Healthcare*, vol. 8, no. 1, p. 34, Feb. 2020, doi: 10.3390/healthcare8010034.

- [37] G. Pahuja, T. N. Nagabhushan, and B. Prasad, “Early detection of parkinson’s disease by using SPECT imaging and biomarkers,” *J. Intell. Syst.*, vol. 29, no. 1, pp. 1329–1344, 2020, doi: 10.1515/jisys-2018-0261.
- [38] H. J. Yoon, K. Cho, W. G. Kim, Y. J. Jeong, J. E. Jeong, and D. Y. Kang, “Heterogeneity by global and textural feature analysis in F-18 FP-CIT brain PET images for diagnosis of Parkinson’s disease,” *Med. (United States)*, vol. 100, no. 35, p. E26961, 2021, doi: 10.1097/MD.00000000000026961.
- [39] T. Pianpanit, S. Lolak, P. Sawangjai, T. Sudhawiyangkul, and T. Wilairasitporn, “Parkinson’s Disease Recognition Using SPECT Image and Interpretable AI: A Tutorial,” *IEEE Sens. J.*, vol. 21, no. 20, pp. 22304–22316, 2021, doi: 10.1109/JSEN.2021.3077949.
- [40] D. Shi *et al.*, “Classification of Parkinson’s disease using a region-of-interest-and resting-state functional magnetic resonance imaging-based radiomics approach,” *Brain Imaging Behav.*, vol. 16, no. 5, pp. 2150–2163, 2022, doi: 10.1007/s11682-022-00685-y.
- [41] T. Mortezaeezadeh, H. Seyedarabi, B. Mahmoudian, and J. P. Islamian, “Imaging modalities in differential diagnosis of Parkinson’s disease: opportunities and challenges,” *Egypt. J. Radiol. Nucl. Med.*, vol. 52, no. 1, 2021, doi: 10.1186/s43055-021-00454-9.
- [42] S. Abbasi, H. Lan, J. Choupan, N. Sheikh-Bahaei, G. Pandey, and B. Varghese, “Deep learning for the harmonization of structural MRI scans: a survey,” *Biomed. Eng. Online*, vol. 23, no. 1, 2024, doi: 10.1186/s12938-024-01280-6.
- [43] X. Cui *et al.*, “Gray Matter Atrophy in Parkinson’s Disease and the Parkinsonian Variant of Multiple System Atrophy: A Combined ROI- and Voxel-Based Morphometric Study,” *Clinics*, vol. 75, no. 3, p. e1505, 2020, doi: 10.6061/clinics/2020/e1505.
- [44] Y. Ya *et al.*, “Machine Learning Models for Diagnosis of Parkinson’s Disease Using Multiple Structural Magnetic Resonance Imaging Features,” *Front. Aging Neurosci.*, vol. 14, no. April, pp. 1–11, 2022, doi: 10.3389/fnagi.2022.808520.

- [45] Z. Karapinar Senturk, “Early diagnosis of Parkinson’s disease using machine learning algorithms,” *Med. Hypotheses*, vol. 138, no. December 2019, p. 109603, May 2020, doi: 10.1016/j.mehy.2020.109603.
- [46] A. Rana, A. Dumka, R. Singh, M. Rashid, N. Ahmad, and M. K. Panda, “An Efficient Machine Learning Approach for Diagnosing Parkinson’s Disease by Utilizing Voice Features,” *Electron.*, vol. 11, no. 22, pp. 1–20, 2022, doi: 10.3390/electronics11223782.
- [47] N. P. George *et al.*, “Integrative Analysis of Metabolome and Proteome in the Cerebrospinal Fluid of Patients with Multiple System Atrophy,” *Cells*, vol. 14, no. 4, 2025, doi: 10.3390/cells14040265.
- [48] H. Tumani *et al.*, “S1-Leitlinie: Lumbalpunktion und Liquordiagnostik,” *DGNeurologie*, vol. 2, no. 6, pp. 456–480, 2019, doi: 10.1007/s42451-019-00126-z.
- [49] S. Lotankar, K. S. Prabhavalkar, and L. K. Bhatt, “Biomarkers for Parkinson’s Disease: Recent Advancement,” *Neurosci. Bull.*, vol. 33, no. 5, pp. 585–597, 2017, doi: 10.1007/s12264-017-0183-5.
- [50] A. Siderowf, “HHS Public Access synuclein seed amplification : a cross-sectional study,” vol. 22, no. 5, pp. 407–417, 2023, doi: 10.1016/S1474-4422(23)00109-6.Assessment.
- [51] M. Nour, U. Senturk, and K. Polat, “Diagnosis and classification of Parkinson’s disease using ensemble learning and 1D-PDCovNN,” *Comput. Biol. Med.*, vol. 161, no. March, p. 107031, 2023, doi: 10.1016/j.combiomed.2023.107031.
- [52] P. R. Magesh, R. D. Myloth, and R. J. Tom, “An Explainable Machine Learning Model for Early Detection of Parkinson’s Disease using LIME on DaTSCAN Imagery,” *Comput. Biol. Med.*, vol. 126, no. July, p. 104041, 2020, doi: 10.1016/j.combiomed.2020.104041.
- [53] M. Nazari *et al.*, “Explainable AI to improve acceptance of convolutional neural networks for automatic classification of dopamine transporter SPECT in the diagnosis of clinically uncertain parkinsonian syndromes,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 49, no. 4, pp. 1176–1186, 2022, doi: 10.1007/s00259-021-05569-9.

- [54] M. Camacho *et al.*, “Explainable classification of Parkinson’s disease using deep learning trained on a large multi-center database of T1-weighted MRI datasets,” *NeuroImage Clin.*, vol. 38, no. April, p. 103405, 2023, doi: 10.1016/j.nicl.2023.103405.
- [55] M. Yang, X. Huang, L. Huang, and G. Cai, “Diagnosis of Parkinson’s disease based on 3D ResNet: The frontal lobe is crucial,” *Biomed. Signal Process. Control*, vol. 85, no. August 2022, p. 104904, Aug. 2023, doi: 10.1016/j.bspc.2023.104904.
- [56] W. Wang, J. Lee, F. Harrou, and Y. Sun, “Early Detection of Parkinson’s Disease Using Deep Learning and Machine Learning,” *IEEE Access*, vol. 8, pp. 147635–147646, 2020, doi: 10.1109/ACCESS.2020.3016062.
- [57] M. Nilashi *et al.*, “Early diagnosis of Parkinson’s disease: A combined method using deep learning and neuro-fuzzy techniques,” *Comput. Biol. Chem.*, vol. 102, no. 1, p. 107788, 2023, doi: 10.1016/j.compbiochem.2022.107788.
- [58] L. Huang, X. Ye, M. Yang, L. Pan, and S. hua Zheng, “MNC-Net: Multi-task graph structure learning based on node clustering for early Parkinson’s disease diagnosis,” *Comput. Biol. Med.*, vol. 152, no. October 2022, p. 106308, Jan. 2023, doi: 10.1016/j.compbioemed.2022.106308.
- [59] G. Pahuja and B. Prasad, “Deep learning architectures for Parkinson’s disease detection by using multi-modal features,” *Comput. Biol. Med.*, vol. 146, no. January, p. 105610, Jul. 2022, doi: 10.1016/j.compbioemed.2022.105610.
- [60] S. Balasubramanian, J. Mandala, T. V. M. Rao, and A. Misra, “RF-ShCNN: A combination of two deep models for tumor detection in brain using MRI,” *Biomed. Signal Process. Control*, vol. 88, no. PA, p. 105656, 2024, doi: 10.1016/j.bspc.2023.105656.
- [61] F. Mohammad and S. Al Ahmadi, “Alzheimer’s Disease Prediction Using Deep Feature Extraction and Optimization,” *Mathematics*, vol. 11, no. 17, pp. 1–17, 2023, doi: 10.3390/math11173712.
- [62] J. Cheng, H. Wang, S. Wei, J. Mei, F. Liu, and G. Zhang, “Alzheimer’s disease prediction algorithm based on de-correlation constraint and multi-modal feature interaction,” *Comput. Biol. Med.*, vol. 170, no. September 2023, p. 108000, 2024, doi: 10.1016/j.compbioemed.2024.108000.

- [63] A. Hamza *et al.*, “D2BOF-COVIDNet: A Framework of Deep Bayesian Optimization and Fusion-Assisted Optimal Deep Features for COVID-19 Classification Using Chest X-ray and MRI Scans,” *Diagnostics*, vol. 13, no. 1, p. 101, Dec. 2022, doi: 10.3390/diagnostics13010101.
- [64] A. Hamza *et al.*, “COVID-19 classification using chest X-ray images based on fusion-assisted deep Bayesian optimization and Grad-CAM visualization,” *Front. Public Heal.*, vol. 10, 2022, doi: 10.3389/fpubh.2022.1046296.
- [65] M. El Haj, C. Boutoleau-Bretonnière, and G. Chapelet, “ChatGPT’s dance with neuropsychological data: A case study in Alzheimer’s disease,” *Ageing Res. Rev.*, vol. 92, no. October, pp. 1–4, 2023, doi: 10.1016/j.arr.2023.102117.
- [66] F. Yang, D. Yan, and Z. Wang, “Large-Scale assessment of ChatGPT’s performance in benign and malignant bone tumors imaging report diagnosis and its potential for clinical applications,” *J. Bone Oncol.*, vol. 44, p. 100525, 2024, doi: 10.1016/j.jbo.2024.100525.
- [67] S. Malik and S. Zaheer, “ChatGPT as an aid for pathological diagnosis of cancer,” *Pathol. Res. Pract.*, vol. 253, no. August 2023, p. 154989, 2024, doi: 10.1016/j.prp.2023.154989.
- [68] A. Bartoli, A. T. May, A. Al-Awadhi, and K. Schaller, “Probing artificial intelligence in neurosurgical training: ChatGPT takes a neurosurgical residents written exam,” *Brain and Spine*, vol. 4, no. August 2023, p. 102715, 2024, doi: 10.1016/j.bas.2023.102715.
- [69] W. A. Campbell, J. F. B. Chick, D. Shin, and M. S. Makary, “Understanding ChatGPT for evidence-based utilization in interventional radiology,” *Clin. Imaging*, vol. 108, no. January, p. 110098, 2024, doi: 10.1016/j.clinimag.2024.110098.
- [70] W. Zhu, W. Geng, L. Huang, X. Qin, Z. Chen, and H. Yan, “Who could and should give exercise prescription: Physicians, exercise and health scientists, fitness trainers, or ChatGPT?,” *J. Sport Heal. Sci.*, vol. 13, no. 3, pp. 368–372, 2024, doi: 10.1016/j.jshs.2024.01.001.
- [71] J. Liu, J. Zheng, X. Cai, D. Wu, and C. Yin, “A descriptive study based on the comparison of ChatGPT and evidence-based neurosurgeons,” *iScience*, vol. 26, no. 9, p. 107590, 2023, doi: 10.1016/j.isci.2023.107590.

- [72] C. Karakas, D. Brock, and A. Lakhotia, “Leveraging ChatGPT in the Pediatric Neurology Clinic: Practical Considerations for Use to Improve Efficiency and Outcomes,” *Pediatr. Neurol.*, vol. 148, pp. 157–163, 2023, doi: 10.1016/j.pediatrneurol.2023.08.035.
- [73] C. Tippareddy, S. Jiang, K. Bera, and N. Ramaiya, “Radiology Reading Room for the Future: Harnessing the Power of Large Language Models Like ChatGPT,” *Curr. Probl. Diagn. Radiol.*, vol. 000, pp. 1–6, 2023, doi: 10.1067/j.cpradiol.2023.08.018.
- [74] D. Hu, B. Liu, X. Zhu, X. Lu, and N. Wu, “Zero-shot information extraction from radiological reports using ChatGPT,” *Int. J. Med. Inform.*, vol. 183, no. December 2023, p. 105321, 2024, doi: 10.1016/j.ijmedinf.2023.105321.
- [75] N. Shifai, R. van Doorn, J. Malvehy, and T. E. Sangers, “Can ChatGPT vision diagnose melanoma? An exploratory diagnostic accuracy study,” *J. Am. Acad. Dermatol.*, vol. 90, no. 5, pp. 1057–1059, 2024, doi: 10.1016/j.jaad.2023.12.062.
- [76] W. Poewe *et al.*, “Parkinson disease,” *Nat. Rev. Dis. Prim.*, vol. 3, no. 1, p. 17013, Mar. 2017, doi: 10.1038/nrdp.2017.13.
- [77] Z. Shu, P. Pang, X. Wu, S. Cui, Y. Xu, and M. Zhang, “An Integrative Nomogram for Identifying Early-Stage Parkinson’s Disease Using Non-motor Symptoms and White Matter-Based Radiomics Biomarkers From Whole-Brain MRI,” *Front. Aging Neurosci.*, vol. 12, no. December, pp. 1–15, Dec. 2020, doi: 10.3389/fnagi.2020.548616.
- [78] F. Salawu, A. Danburam, and A. Olokoba, “Non-Motor Symptoms of Parkinson’s Disease: Diagnosis and Management,” *Niger. J. Med.*, vol. 19, no. 2, pp. 126–131, Jul. 2010, doi: 10.4314/njm.v19i2.56496.
- [79] B. Garcia Santa Cruz, A. Husch, and F. Hertel, “Machine learning models for diagnosis and prognosis of Parkinson’s disease using brain imaging: general overview, main challenges, and future directions,” *Front. Aging Neurosci.*, vol. 15, no. M1, 2023, doi: 10.3389/fnagi.2023.1216163.
- [80] X. Li, P. S. Morgan, J. Ashburner, J. Smith, and C. Rorden, “The first step for neuroimaging data analysis: DICOM to NIfTI conversion,” *J. Neurosci. Methods*, vol. 264, pp. 47–56, 2016, doi: 10.1016/j.jneumeth.2016.03.001.

- [81] S. M. Smith, “Fast robust automated brain extraction,” *Hum. Brain Mapp.*, vol. 17, no. 3, pp. 143–155, Nov. 2002, doi: 10.1002/hbm.10062.
- [82] Y. Xiao *et al.*, “A dataset of multi-contrast population-averaged brain MRI atlases of a Parkinson’s disease cohort,” *Data Br.*, vol. 12, pp. 370–379, Jun. 2017, doi: 10.1016/j.dib.2017.04.013.
- [83] N. J. Tustison *et al.*, “N4ITK: Improved N3 Bias Correction,” *IEEE Trans. Med. Imaging*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010, doi: 10.1109/TMI.2010.2046908.
- [84] L. J. Isaksson *et al.*, “Effects of MRI image normalization techniques in prostate cancer radiomics,” *Phys. Medica*, vol. 71, no. February, pp. 7–13, Mar. 2020, doi: 10.1016/j.ejmp.2020.02.007.
- [85] J. Li *et al.*, “Cortical and subcortical morphological alterations in motor subtypes of Parkinson’s disease,” *npj Park. Dis.*, vol. 8, no. 1, p. 167, Dec. 2022, doi: 10.1038/s41531-022-00435-3.
- [86] S. Sivaranjini and C. M. Sujatha, “Morphological analysis of subcortical structures for assessment of cognitive dysfunction in Parkinson’s disease using multi-atlas based segmentation,” *Cogn. Neurodyn.*, vol. 15, no. 5, pp. 835–845, Oct. 2021, doi: 10.1007/s11571-021-09671-4.
- [87] C. J. Park *et al.*, “An interpretable multiparametric radiomics model of basal ganglia to predict dementia conversion in Parkinson’s disease,” *npj Park. Dis.*, vol. 9, no. 1, p. 127, Aug. 2023, doi: 10.1038/s41531-023-00566-1.
- [88] J.-H. Deng, H.-W. Zhang, X.-L. Liu, H.-Z. Deng, and F. Lin, “Morphological changes in Parkinson’s disease based on magnetic resonance imaging: A mini-review of subcortical structures segmentation and shape analysis,” *World J. Psychiatry*, vol. 12, no. 12, pp. 1356–1366, Dec. 2022, doi: 10.5498/wjp.v12.i12.1356.
- [89] Y. Wu *et al.*, “Use of radiomic features and support vector machine to distinguish Parkinson’s disease cases from normal controls,” *Ann. Transl. Med.*, vol. 7, no. 23, pp. 773–773, Dec. 2019, doi: 10.21037/atm.2019.11.26.

- [90] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, no. 2, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
- [91] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Section 2, pp. 4766–4775, 2017.
- [92] M. Ribeiro, S. Singh, and C. Guestrin, ““Why Should I Trust You?”: Explaining the Predictions of Any Classifier,” in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2016, pp. 97–101. doi: 10.18653/v1/N16-3020.
- [93] S. J. Chung, H. S. Lee, H. S. Yoo, Y. H. Lee, P. H. Lee, and Y. H. Sohn, “Patterns of striatal dopamine depletion in early Parkinson disease: Prognostic relevance,” *Neurology*, vol. 95, no. 3, pp. E280–E290, 2020, doi: 10.1212/WNL.0000000000009878.
- [94] D. Molina *et al.*, “Influence of gray level and space discretization on brain tumor heterogeneity measures obtained from magnetic resonance images,” *Comput. Biol. Med.*, vol. 78, pp. 49–57, 2016, doi: 10.1016/j.combiomed.2016.09.011.
- [95] N. J. Tustison and J. Gee, “Run-Length Matrices For Texture Analysis,” *Insight J.*, vol. 1, no. 2, pp. 1–6, 2022, doi: 10.54294/ex0itu.
- [96] C. E. Clarke, “Parkinson’s disease,” *Br. Med. J.*, vol. 335, no. 7617, pp. 441–445, 2007, doi: 10.1136/bmj.39289.437454.AD.
- [97] A. W. Willis *et al.*, “Incidence of Parkinson disease in North America,” *npj Park. Dis.*, vol. 8, no. 1, 2022, doi: 10.1038/s41531-022-00410-y.
- [98] P. Surathi, K. Jhunjhunwala, R. Yadav, and P. K. Pal, “Research in Parkinson’s disease in India: A review,” *Ann. Indian Acad. Neurol.*, vol. 19, no. 1, pp. 9–20, 2016, doi: 10.4103/0972-2327.167713.
- [99] M. Shalaby, N. A. Belal, and Y. Omar, “Classification of Parkinson’s Disease,” vol. 2021, no. 18, pp. 1587–1593, 2021.

- [100] R. Prashanth, S. Dutta Roy, P. K. Mandal, and S. Ghosh, “Automatic classification and prediction models for early Parkinson’s disease diagnosis from SPECT imaging,” *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3333–3342, Jun. 2014, doi: 10.1016/j.eswa.2013.11.031.
- [101] S. M. Abdullah *et al.*, “Deep Transfer Learning Based Parkinson’s Disease Detection Using Optimized Feature Selection,” *IEEE Access*, vol. 11, no. January, pp. 3511–3524, 2023, doi: 10.1109/ACCESS.2023.3233969.
- [102] N. U. Islam, R. Khanam, and A. Kumar, “Using 3D CNN for classification of Parkinson’s disease from resting-state fMRI data,” *J. Eng. Appl. Sci.*, vol. 70, no. 1, pp. 1–13, 2023, doi: 10.1186/s44147-023-00236-2.
- [103] F. Ullah *et al.*, “Evolutionary Model for Brain Cancer-Grading and Classification,” *IEEE Access*, vol. 11, no. October, pp. 126182–126194, 2023, doi: 10.1109/ACCESS.2023.3330919.
- [104] N. Aggarwal, B. S. Saini, and S. Gupta, “A deep 1-D CNN learning approach with data augmentation for classification of Parkinson’s disease and scans without evidence of dopamine deficit (SWEDD),” *Biomed. Signal Process. Control*, vol. 91, no. December 2023, p. 106008, 2024, doi: 10.1016/j.bspc.2024.106008.
- [105] D. Kollias, A. Tagaris, A. Stafylopatis, S. Kollias, and G. Tagaris, “Deep neural architectures for prediction in healthcare,” *Complex Intell. Syst.*, vol. 4, no. 2, pp. 119–131, 2018, doi: 10.1007/s40747-017-0064-6.
- [106] S. Shinde *et al.*, “Predictive markers for Parkinson’s disease using deep neural nets on neuromelanin sensitive MRI,” *NeuroImage Clin.*, vol. 22, no. January, p. 101748, 2019, doi: 10.1016/j.nicl.2019.101748.
- [107] A. Bhan, S. Kapoor, and M. Gulati, “Diagnosing Parkinson’s disease in Early Stages using Image Enhancement, ROI Extraction and Deep Learning Algorithms,” in *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, IEEE, Apr. 2021, pp. 521–525. doi: 10.1109/ICIEM51511.2021.9445381.
- [108] S. Mirjalili and A. Lewis, “The Whale Optimization Algorithm,” *Adv. Eng. Softw.*, vol. 95, pp. 51–67, 2016, doi: 10.1016/j.advengsoft.2016.01.008.

- [109] E. Emary *et al.*, “A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases,” *J. Supercomput.*, vol. 21, no. 11, pp. 11797–11826, 2013, doi: 10.1016/j.neucom.2016.03.101.
- [110] E. Emary, H. M. Zawbaa, and A. E. Hassanien, “Binary ant lion approaches for feature selection,” *Neurocomputing*, vol. 213, pp. 54–65, 2016, doi: 10.1016/j.neucom.2016.03.101.
- [111] A. B. R. Leite Silva *et al.*, “Premotor, nonmotor and motor symptoms of Parkinson’s Disease: A new clinical state of the art,” *Ageing Res. Rev.*, vol. 84, no. December 2022, 2023, doi: 10.1016/j.arr.2022.101834.
- [112] P. Macías-García *et al.*, “Neuropsychiatric Symptoms in Clinically Defined Parkinson’s Disease: An Updated Review of Literature,” *Behav. Neurol.*, vol. 2022, no. Table 1, 2022, doi: 10.1155/2022/1213393.
- [113] M. Junaid, S. Ali, F. Eid, S. El-Sappagh, and T. Abuhmed, “Explainable machine learning models based on multimodal time-series data for the early detection of Parkinson’s disease,” *Comput. Methods Programs Biomed.*, vol. 234, p. 107495, Jun. 2023, doi: 10.1016/j.cmpb.2023.107495.
- [114] E. Porter, A.-A. Roussakis, N. P. Lao-Kaim, and P. Piccini, “Multimodal dopamine transporter (DAT) imaging and magnetic resonance imaging (MRI) to characterise early Parkinson’s disease,” *Parkinsonism Relat. Disord.*, vol. 79, no. August, pp. 26–33, Oct. 2020, doi: 10.1016/j.parkreldis.2020.08.010.
- [115] V. Skaramagkas, A. Pentari, Z. Kefalopoulou, and M. Tsiknakis, “Multi-Modal Deep Learning Diagnosis of Parkinson’s Disease - A Systematic Review,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2399–2423, 2023, doi: 10.1109/TNSRE.2023.3277749.
- [116] J. Yang, H. B. Li, and D. Wei, “The impact of ChatGPT and LLMs on medical imaging stakeholders: Perspectives and use cases,” *Meta-Radiology*, vol. 1, no. 1, p. 100007, 2023, doi: 10.1016/j.metrad.2023.100007.
- [117] V. Sorin *et al.*, “Large language model (ChatGPT) as a support tool for breast tumor board,” *npj Breast Cancer*, vol. 9, no. 1, pp. 1–4, 2023, doi: 10.1038/s41523-023-00557-8.

- [118] Y. Xiang *et al.*, “Estimating the sequence of biomarker changes in Parkinson’s disease,” *Park. Relat. Disord.*, vol. 118, no. July 2023, p. 105939, 2024, doi: 10.1016/j.parkreldis.2023.105939.
- [119] G. Briganti, “How ChatGPT works: a mini review.,” *Eur. Arch. oto-rhino-laryngology Off. J. Eur. Fed. Oto-Rhino-Laryngological Soc. Affil. with Ger. Soc. Oto-Rhino-Laryngology - Head Neck Surg.*, vol. 281, no. 3, pp. 1565–1569, Mar. 2024, doi: 10.1007/s00405-023-08337-7.
- [120] M. Thelwall, “ChatGPT for complex text evaluation tasks,” *J. Assoc. Inf. Sci. Technol.*, no. November, pp. 1–4, 2024, doi: 10.1002/asi.24966.
- [121] Y. Zhou *et al.*, “Cloud-magnetic resonance imaging system: In the era of 6G and artificial intelligence,” *Magn. Reson. Lett.*, no. xxxx, p. 200138, 2024, doi: 10.1016/j.mrl.2024.200138.
- [122] R. Prashanth, S. Dutta Roy, P. K. Mandal, and S. Ghosh, “High-Accuracy Detection of Early Parkinson’s Disease through Multimodal Features and Machine Learning,” *Int. J. Med. Inform.*, vol. 90, pp. 13–21, 2016, doi: 10.1016/j.ijmedinf.2016.03.001.

## APPENDIX A: LIST OF ABBREVIATIONS

1D-PDCov NN	One-Dimensional Partial Distance Covariance Neural Network
ACO	Ant Colony Optimization
AD	Alzheimer's Disease
AI	Artificial intelligence
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AUC	Area Under the Curve
$\text{A}\beta_{1-42}$	Amyloid Beta <sub>1-42</sub>
BERT	Bidirectional Encoder Representations from Transformers
BN	Batch Normalization
BioBERT	Biomedical BERT
CCA	Canonical Correlation Analysis
CCDFs	Clinical Decision Support System
CNN	Convolutional Neural Network
CSF	Cerebrospinal Fluid
D2BOF	Deep Bayesian-Optimized Framework
DBN	Deep Belief Network
DBS	Deep Brain Stimulation
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DLBLSTM	Densely Linked Bidirectional Long Short-Term Memory

DT	Decision Tree
DTI	Diffusion Tensor Imaging
DWI	Diffusion-Weighted Imaging
DaTscan	Dopamine Transporter Imaging
EEG	Electroencephalogram
EMG	Electromyography
XT	Extra trees
FLAIR	Fluid-Attenuated Inversion Recovery
FN	False Negative
FP	False Positive
FSL	FMRIB Software Library
GA	Genetic Algorithm
GAN	Generative Adversarial Network
GB	Gradient Boosting
GLCM	Gray Level Co-occurrence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
ChatGPT	Chat Generative Pre-trained Transformer
Gpe	Globus Pallidus Externus
GPi	Globus Pallidus Internus
GSA	Gravitational Search Algorithm
Grad-CAM	Gradient-weighted Class Activation Mapping
H & Y scale	Hoehn and Yahr Scale
HC	Healthy control
IR	Inversion Recovery

JSON	JavaScript Object Notation
KNN	K-Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
LIME	Local Interpretable Model-Agnostic Explanations
LLM	Large Language Model
LRP	Layer-wise Relevance Propagation
LRRK2	Leucine-Rich Repeat Kinase 2
ML	Machine Learning
MLR	Multivariate Logistic Regression
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
MSA	Multiple System Atrophy
MoCA	Montreal Cognitive Assessment
NGTDM	Neighbouring Gray Tone Difference Matrix
NIfTI	Neuroimaging Informatics Technology Initiative
PARK2	Parkin Gene
PD	Parkinson's disease
PET	Positron Emission Tomography
PPMI	Parkinson's Progression Markers Initiative
PSO	Particle Swarm Optimization
PSP	Progressive Supranuclear Palsy
RBD	Rapid Eye Movement Sleep Behavior Disorder
RFE	Recursive Feature Elimination
RLHF	Reinforcement Learning with Human Feedback
RNN	Recurrent Neural Network

ROI	Region of Interest
ROS	Reactive Oxygen Species
ROC	Receiver Operating Characteristic
ReLU	Rectified Linear Unit
SBR	Specific Binding Ratio
SHAP	Shapley Additive Explanations
SMOTE	Synthetic Minority Over-sampling Technique
SN	Substantia Nigra
SPECT	Single Photon Emission Computed Tomography
SVM	Support Vector Machine
TE	Echo Time
TI	Inversion Time
TN	True Negative
TP	True Positive
TR	Repetition Time
UPDRS	Unified Parkinson's Disease Rating Scale
WOA	Whale Optimization Algorithm
XAI	Explainable Artificial Intelligence
fMRI	Functional Magnetic Resonance Imaging
p-tau	Phosphorylated Tau
t-tau	Total Tau

## APPENDIX B: LIST OF PUBLICATIONS

1. Priyadharshini, S., Ramkumar, K., Venkatesh, S., Narasimhan, K., & Adalarasu, K. (2023). An Overview of Interpretability Techniques for Explainable Artificial Intelligence (XAI) in Deep Learning-based Medical Image Analysis. In Proceedings of the 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Vol. 1, pp. 175–182. IEEE. <https://doi.org/10.1109/ICACCS57279.2023.10113001>. [Conference Proceedings, Scopus Indexed]
  
2. Priyadharshini, S., Ramkumar, K., Vairavasundaram, S., Narasimhan, K., Venkatesh, S., Amirtharajan, R., & Madhavasarma, P. (2024). A Comprehensive Framework for Parkinson's Disease Diagnosis Using Explainable Artificial Intelligence Empowered Machine Learning Techniques. Alexandria Engineering Journal, 107, 568–582. <https://doi.org/10.1016/j.aej.2024.07.106>. [SCIE, Impact Factor: 6.2]
  
3. Priyadharshini, S., Ramkumar, K., Vairavasundaram, S., Narasimhan, K., Venkatesh, S., Madhavasarma, P., & Adalarasu, K. (2024). Bio-inspired Feature Selection for Early Diagnosis of Parkinson's Disease Through Optimization of Deep 3D Nested Learning. Scientific Reports, 14, 1–17. <https://doi.org/10.1038/s41598-024-74405-5>. [SCIE, Impact Factor: 3.8]
  
4. Priyadharshini, S., Narasimhan, K., Ramkumar, K., & Venkatesh, S. (2025). Voxel-based Analysis and Advanced Techniques for MRI Scan Classification in Early Diagnosis of Parkinson's Disease. In S. Dey, V. Kumar, D.K. Pratihar, S. Islam, & V.P. Singh (Eds.), Advancing Healthcare through Decision Intelligence (pp. 175–194). Academic Press, Elsevier. <https://doi.org/10.1016/B978-0-443-26480-1.00003-5>. [Book Chapter, Scopus Indexed]
  
5. Priyadharshini, S., Ramkumar, K., Narasimhan, K., & Venkatesh, S. A Web-based AI-driven Framework Integrating Multi-modal Data using CNN and LLMs for Parkinson's Disease Diagnosis. Submitted to Scientific Reports. [ SCIE, Impact Factor: 3.8]

