



A Comprehensive framework for Parkinson's disease diagnosis using explainable artificial intelligence empowered machine learning techniques



S. Priyadarshini^a, K. Ramkumar^b, Subramaniyaswamy Vairavasundaram^{c,*}, K. Narasimhan^a, S. Venkatesh^a, Rengarajan Amirtharajan^a, Ketan Kotcha^{d,*}

^a School of Electrical and Electronics Engineering, SASTRA Deemed University, Thanjavur, India

^b School of Computing, SASTRA Deemed University, Thanjavur, India

^c School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

^d Symbiosis Centre for Applied Artificial Intelligence, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India

ARTICLE INFO

Keywords:

Parkinson's disease
Gradient boosting Algorithm
Magnetic resonance imaging analysis
Explainable artificial intelligence
Early detection

ABSTRACT

Parkinson's disease (PD) is the second most prevalent neurological disorder, predominantly affecting older people. With no existing cure, the early detection of PD, where symptoms are not entirely evident but indicative of the disease's onset, is critical. This study aims to design and develop AI-based diagnostic methods that can detect these early signs of PD with high accuracy, thereby facilitating more effective disease management. This study focuses on developing a method that not only identifies PD at an early stage but also provides clinicians with a tool to interpret the decisions taken by the AI models to avoid misdiagnosis. In this study, a T2-weighted 3D Magnetic resonance imaging (MRI) dataset is used to analyze detailed morphological, textural, and structural changes. The MRI scans are pre-processed using brain extraction, image registration, bias correction, normalization, and segmentation processes. Upon segmentation, feature extraction was applied to the segmented subcortical regions using radiomics tools, resulting in the extraction of 107 features. The top 20 features were selected through Pearson's correlation, recursive feature elimination, and a ranking model, which are responsible for the ML model's class prediction. Statistical validation of these features was also performed using Analysis of Variance (ANOVA), pairwise t-tests, and Kruskal-Wallis H-tests to ensure that the identified 20 features were dominant for the prediction. Based on the identified features, several Machine Learning (ML) models were used to identify the best classifier for the provided real-world MRI scans. The Gradient Boosting (GB) algorithm achieved better prediction accuracy among the compared models. Incorporating the Synthetic Minority Over-sampling Technique (SMOTE) to address data imbalances significantly improved the model's performance, boosting accuracy to 96.8% from 87%. Further, multiple Explainable Artificial Intelligence (XAI) techniques were deployed to enhance the transparency and interpretability of the models. These techniques provide insights into how each identified feature influences predictions by the classifier, assisting clinicians in making trustworthy decisions when planning diagnosis and treatment measures.

1. Introduction

PD is a chronic movement disorder characterized by the deterioration of the motor system and coordination because of the degeneration of dopamine-producing neurons in the brain. It manifests in both motor and non-motor symptoms such as tremors, rigidity, slowed movements, and psychological disturbances, including depression, anxiety, sleep issues, and cognitive decline [1–3]. Previous research points out that

reduced iron levels in the substantia nigra (SN) could be used as potential biomarkers for early diagnosis of certain conditions, likely PD [4, 5]. Traditionally, PD was diagnosed using symptom observation, which maps the spread of the disease in the patient. Even though this method of identification of PD is valid, it can be identified only after the patient is affected by the PD. Since there is no permanent cure for PD, there is a need to detect the progression of PD early.

Over the past two decades, modern imaging technologies such as

* Corresponding authors.

E-mail addresses: priyadarshini@eie.sastra.ac.in (S. Priyadarshini), ramkumar@eie.sastra.edu (K. Ramkumar), subramaniyaswamy.v@vit.ac.in (S. Vairavasundaram), knr@ece.sastra.edu (K. Narasimhan), esvee@eie.sastra.edu (S. Venkatesh), amir@ece.sastra.edu (R. Amirtharajan), director@sitpune.edu.in (K. Kotcha).

MRI, Positron Emission Tomography (PET), and Single-Photon Emission Computed Tomography (SPECT) have enhanced dopamine activity assessment and are used for identifying PD at the early stages [6,7]. Among all, MRI is pivotal in confirming and monitoring PD because of its non-invasive nature, high-resolution capabilities, and absence of radiation, making it safer and more accurate mapping of the disease progression. T2-weighted MRI scans are crucial for identifying PD indicators like structural brain changes and iron accumulation due to their sensitivity to magnetic susceptibility effects and superior soft tissue contrast, whereas T1-weighted and Fluid-Attenuated Inversion Recovery (FLAIR) scans are less effective for these specific purposes [8,9]. MRI data-driven decision-making techniques have several drawbacks that impact PD early diagnosis, even though they are useful in detecting disease development in its early phases. One such limitation is that MRI scan data might exhibit slower and milder features, which, though used to identify disease progression, may not provide a reliable classification framework for early detection of PD. Additionally, these methods can be

unaffordable for large segments of the population due to high costs. Furthermore, the effective mapping of PD from MRI relies on the expertise of the physician, and there is a risk of misdiagnosis in the initial stages due to human error.

The development of PD has been successfully mapped through the analysis of patterns and relationships seen in MRI scans using data-driven models like machine learning (ML) and deep learning (DL) [10, 11]. These methods effectively identify disease progression in its early stages, even with the subtle and milder features available in MRI scans, and eliminate the human error inherent in clinician-based manual evaluations. However, while incorporating ML/DL algorithms for early disease identification is effective, the models require a sufficient amount of data for training to classify efficiently. The complexity of these high-accuracy models often results in a lack of transparency in their evaluation, making it difficult for clinicians to assess patients and formulate treatment plans [12,13].

This study presents a novel approach that integrates XAI with ML

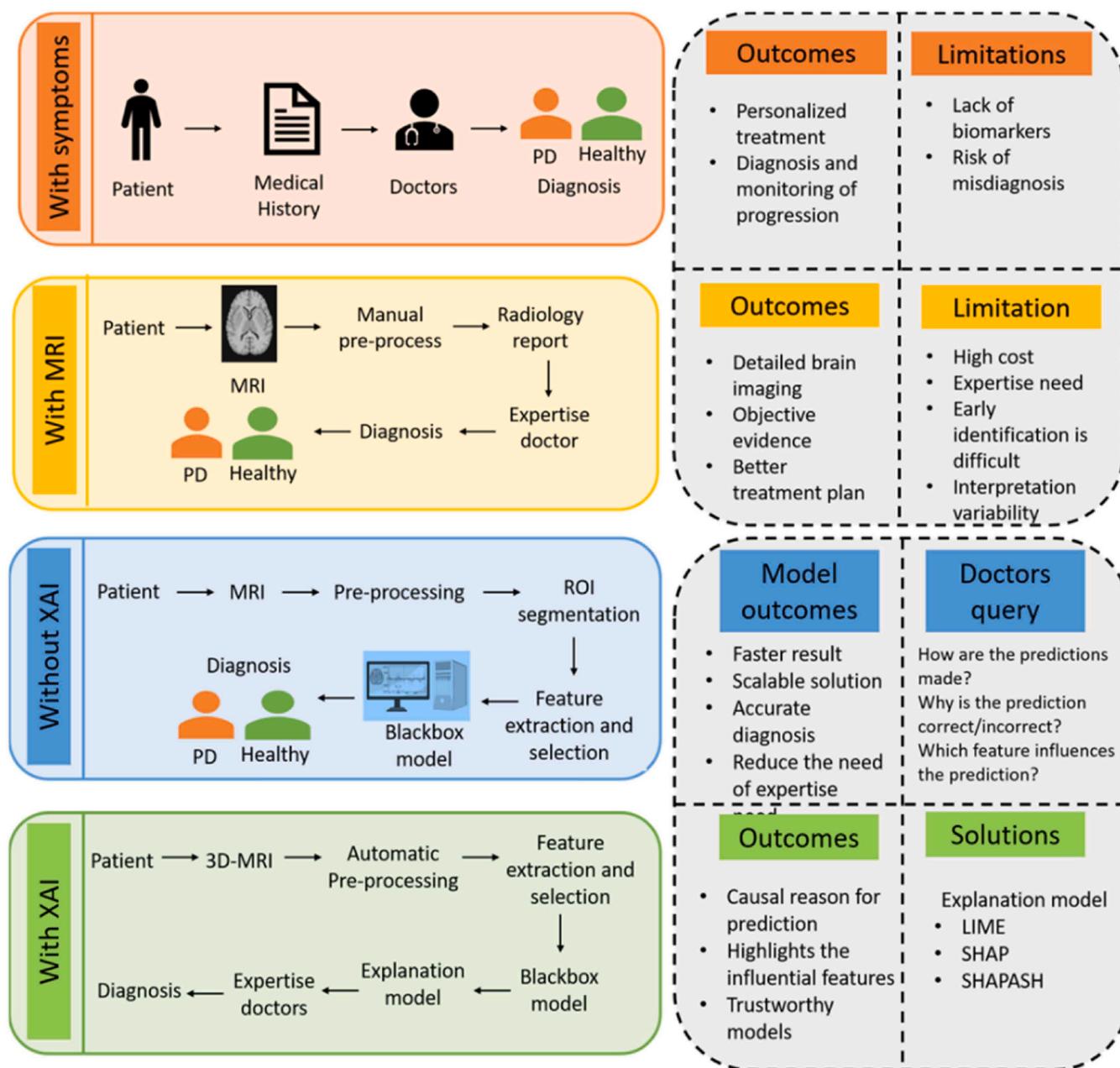


Fig. 1. Process flow of the proposed framework for early detection and monitoring of PD using XAI and ML techniques.

techniques, aiming to increase the transparency and interpretability of these models. This integration is intended to improve clinical acceptance and trust in the automated diagnostic systems. By illuminating the decision-making processes of AI models, XAI allows clinicians to understand the rationale behind the 'black-box' decisions, highlighting critical diagnostic features of PD and deepening our understanding of its underlying mechanisms, which traditional ML algorithm decisions do not provide. These insights are crucial for developing personalized treatment plans that account for individual risk factors, significantly enhancing treatment outcomes. This research primarily focuses on a detailed analysis of structural, morphological, texture, and statistical radiomics features from MRI data across eight subcortical regions to improve PD detection. Incorporating XAI into the proposed framework will support clinicians in making more informed decisions, reducing the risk of misdiagnosis and enabling timely interventions in the diagnostic process. The detailed overview of the proposed framework, along with the limitations of the previous methodologies, are illustrated in Fig. 1. The key contributions of this research article include:

1. Development of a novel framework combining 3D MRI imaging and GB algorithm to enhance the early detection of PD using real-world data, significantly improving diagnostic accuracy.
2. Implementation of multiple XAI methodologies to increase the transparency and trustworthiness of AI-driven diagnostics in PD prediction, facilitating wider clinical acceptance.
3. Radiomics features were used to extract 107 features from subcortical brain regions, identifying 20 critical key features using an optimal feature-selecting strategy.
4. Enhanced the robustness of ML model prediction accuracy by effectively addressing data imbalances with the SMOTE, significantly improving model performance and elevating diagnostic accuracy from 87 % to 96.8 % post-SMOTE application.
5. The proposed framework helped to identify PD's prodromal stage early and created opportunities to plan personalized treatment approaches that could improve patients' quality of life.

The rest of the paper is organized as follows: Section 2 provides a brief overview of the existing literature connected with this research work with an emphasis on various imaging techniques, machine learning algorithms and explainable AI techniques. Section 3 provides the methodology involved in developing the proposed system, elaborating the processing methods applied and the feature selection strategy employed. Section 4 comprehensively covers the results obtained from evaluating the performance of six ML model classifiers and also presents the effect of applying three XAI approaches in providing deeper insights about the causal features involved in the predictions. Section 5 presents a detailed discussion of the proposed XAI-based framework and its performance compared with other approaches. Finally, Section 6 concludes the paper with a discussion on future scope.

2. Related works

This section provides the literature review of the study, which facilitated the development of the proposed framework. PD is a progressive neurological disorder characterized by motor and non-motor symptoms. The application of artificial intelligence (AI) is widely integrated into PD medical diagnostics because its usage contributes to the increase in accuracy and early diagnosis, which is mandatory for further management and treatment of the disease. AI must be integrated into traditional diagnostic techniques in order to increase diagnostic precision and reliability because these techniques frequently rely on clinical judgments and are prone to subjective interpretation. The diagnosis of Parkinson's disease (PD) involves many difficulties in traditional clinical settings. In [14], the author found that 15 % of patients diagnosed with PD did not meet the strict clinical criteria, while 20 % of actual PD cases went undiagnosed, highlighting the diagnostic difficulties and the need

for more rigorous criteria. Similarly, in [15], authors discuss how PD is misdiagnosed since its symptoms might be mistaken for those of other neurodegenerative diseases, and there are no reliable biomarkers for the disease. These difficulties highlight how crucial it is to regularly reevaluate and interact with experts in order to increase diagnostic precision.

Detecting PD involves analyzing various data modalities such as neuroimaging data (such as MRI, PET, and SPECT scans), biomarkers (like dopamine transporter levels), and proteins (alpha-synuclein, tau). In [16], the authors conclude that biomarkers present useful ways of diagnosing early PD, assessing progressive PD, and differentiating PD from other disorders, including Lewy bodies (DLB). Neuroimaging techniques such as structural MRI and neuromelanin-sensitive MRI distinguish PD from DLB based on the presence of brain atrophy or differences in neuromelanin content. DAT SPECT helps to confirm PD diagnoses with great accuracy by assessing the utilization of dopamine transporter and F-DOPA PET to assess the correlation with the number of nigral cells and motor symptoms. For instance, in [17], the author applies the approach of machine learning on MDVP audio data on the early detection of PD, which gave high accuracy to the selected Random Forest (RF) classifiers. An ML model that emphasizes the importance of interpretability in early diagnosis is presented in [18], utilizing dopamine transporter (DaTSCAN) images. In order to differentiate between PD patients and healthy persons, texture analysis and machine learning are used in reference [19] to investigate the use of F-18 FP-CIT PET scans for PD diagnosis. Reference [20] highlights the use of deep convolutional neural networks (DCNN) for PD recognition from SPECT images, emphasizing model interpretability. In [21], textural features from diffusion MRI are used to detect PD, achieving high accuracy in classification, while reference [22] discusses a comprehensive AI approach for advanced PD detection using clinical assessments and neuroimaging samples.

ML and DL techniques effectively leverage MRI data for accurate and early diagnosis of PD. These techniques identify distinct patterns and biomarkers that conventional approaches fail to detect, and these strategies improve the accuracy of diagnosis. Integrating AI with MRI aids in better disease monitoring and treatment planning. Because of its detailed brain imaging and great spatial resolution, MRI data is better for PD identification. It examines morphological and structural changes in the brain, including changes in texture, shape, volume, and structure. Voxel-based morphometry (VBM) is a specific neuroimaging analysis technique that utilizes MRI scans to study differences in brain anatomy. By applying VBM to high-resolution MRI data, researchers can measure and analyze variations in grey matter concentration and brain morphology, allowing for the detection of structural changes associated with neurological conditions.

In [23], VBM highlights grey matter reductions in PD patients with wearing-off symptoms, associating cortical atrophy with motor symptoms. These findings enhance the understanding of neurodegeneration in PD and its relation to cognitive decline and motor impairment. In [24], the author discusses the longitudinal volumetric and shape changes in subcortical nuclei in PD patients using advanced shape analysis techniques for MRI scans. In [25], the integration of clinical and radiomics features to identify and predict PD subtypes with high accuracy is highlighted. References [26,27] explore the use of Bayesian network classifiers and fuzzy k-nearest neighbor (KNN) approaches for PD prediction, emphasizing the effectiveness of these methods in handling complex, high-dimensional data. Reference [28] investigates texture-based markers from structural MRI correlating with motor handicap in PD, providing valuable biomarkers for disease progression. In [29–31], researchers focus on the advancements and challenges in radiomics, highlighting its potential to enhance personalized cancer treatment and its applications in differentiating PD from other Parkinsonian syndromes.

AI has significantly improved the accuracy of PD diagnosis, but challenges remain in handling and interpreting high-quality brain

imaging data. These challenges can delay research and limit clinical application, as well as the lack of transparency in AI-driven methods, as ML and DL models are ‘black boxes,’ limiting their acceptance by medical professionals. With transparency, medical professionals can trust AI-driven diagnostic tools. This drives the shift towards XAI and makes AI model decisions comprehensible, enhancing trust and acceptance, especially in healthcare. In [32], the author discusses the need for interpretability in AI models, emphasizing the importance of making AI decisions transparent and understandable. In [33], an explainable deep learning model for classifying PD using T1-weighted MRI datasets highlights the use of saliency maps to identify crucial brain regions influencing classification. Reference [34] investigates the use of layer-wise relevance propagation (LRP) to provide interpretability for CNN-based classification of DAT-SPECT images, demonstrating the importance of XAI in enhancing clinical acceptance and trust in automated diagnostic tools.

By improving interpretability, XAI makes AI-driven insights more accessible and trustworthy for medical professionals. Our proposed study utilizing radiomics features with XAI enhances the understanding of complex data, providing clear and actionable insights. This study leverages a novel ML-based classification framework that integrates XAI techniques, aiming to bridge the gap between technical accuracy and clinical usability, making AI-driven insights more accessible for medical professionals. A comparison analysis of neuroimaging techniques for PD diagnosis is shown in Table 1.

3. Proposed method

3.1. Data collection

The study utilizes MRI scans from the Parkinson’s Progression Markers Initiative (PPMI) database, which is accessible at (www.ppmi-info.org/data). To proceed with the analysis, the images are initially downloaded in Digital Imaging and Communications in Medicine (DICOM) format for each patient and subsequently converted into Neuroimaging Informatics Technology Initiative (NIFTI) format [36]. This conversion facilitates the 3D analysis necessary for our study, providing views of axial, sagittal, and coronal planes. The MRI scans were obtained using a SIEMENS scanner located in Munich, Germany. The dataset includes MRI scans of 500 patients categorized into three

groups: 180 with PD, 160 with prodromal symptoms, and 160 HC. Details on the MRI scan parameters are provided in Table 2.

3.2. Data preprocessing

The MRI processing pipeline for our study involves several advanced computational techniques to prepare the scans for detailed analysis, as illustrated in Fig. 2. The process begins by initializing with NIFTI files for each patient, aiming to produce bias-corrected, registered, and segmented images.

3.2.1. Brain extraction

Each NIFTI file is subjected to the process of brain region extraction using the Robust Brain Extraction (ROBEX) algorithm. This algorithm generates a binary mask that distinguishes brain (1) from non-brain tissue (0) by iteratively refining the boundary between brain and non-brain regions using a combination of morphological operations and probabilistic models [37].

Table 2
Parameter for imaging protocols.

Imaging protocol	Values
Modality	MRI
Research Group	PD, Control, Prodromal
Visit	Baseline (BL)
Acquisition Plane, type	SAGITTAL(3D)
Field Strength	3.0 tesla
Angle	120.0 degree
Manufacturer	SIEMENS (SAG 3D FLAIR)
Matrix	X=256–512 pixels Y=256–512 pixels Z=20–60 pixels
Pixel spacing	X=0.5–1.5 mm Y=0.5–1.5 mm Z=2–5 mm
Pulse sequence	Spin Echo (SE)/ Inversion Recovery (IR)
Slice thickness	1.0 mm
TE/TI/TR	120–500 ms/1800–2200 ms/2000–5000 ms
Weighting	T2

Table 1

Summary of key findings from recent studies on neuroimaging techniques for PD diagnosis using Artificial Intelligence and Machine Learning.

Reference	Year	Dataset details	Data	Methodology	Region of Interest (ROI)	Findings
[31]	2021	65-PD 61-APS 30-MSA 75-HC	MRI Scan	RF classifier, Radiomics feature extraction, 5-fold cross-validation, 80 % split for train, 20 % split for test	Thalamus, Caudate, Putamen, SN, cerebellum, and brain stem	PD vs HC – 70 % accuracy PD vs APS – 92 % accuracy PD vs APS vs HC – 96 % accuracy
[35]	2020	356-HC, 355- PD	dMRI scan	Custom network, 10-fold cross-validation, cross-entropy loss, Adam optimizer	12 ROI (six on the hemisphere of the brain, dorsal premotor cortex, supplementary motor area	97.8 % of accuracy
[21]	2022	123-PD 90-HC	MRI data	Radiomics feature extraction, Two-sample t-test, RFE with SVM for feature selection and 5-fold cross-validation	210 cortical and 36 subcortical regions in the cerebellum	78.07 % -accuracy 0.859-AUC 78.80 %- Sensitivity 76.08 %-Specificity
[28]	2021	32-HC 39-early stage of PD 37-late stage of PD	MRI data	Statistical analysis and first & second-order texture feature extraction (ROI and VOI are compared)	SN, striatum, thalamus and subthalamic nucleus	Texture features associated with clinical scores were significant in PD. Volume changes in the PD group were not significantly different.
[24]	2024	32- HC 34-PD	MRI	Volume and shape change analysis is performed between subcortical and clinical symptoms.	Subcortical region (putamen and pallidum)	PD is significant with left pallidum and putamen. Left putamen is negatively correlated with clinical symptoms.
[29]	2023	36-HC 47-PD	MRI	SVM and RF, 10-fold cross-validation, radiomics features, custom model, 75 % of train, 25 % of test	SN region	83.33 % of accuracy

Abbreviations: APS- Atypical Parkinsonian syndromes; MSA- Multiple System Atrophy, HC-Healthy control; dmri-diffusion Magnetic resonance images, RFE- Recursive Feature elimination, VOI- Volume of Interest, SVM- Support Vector Machine; AUC- Area under the curve.

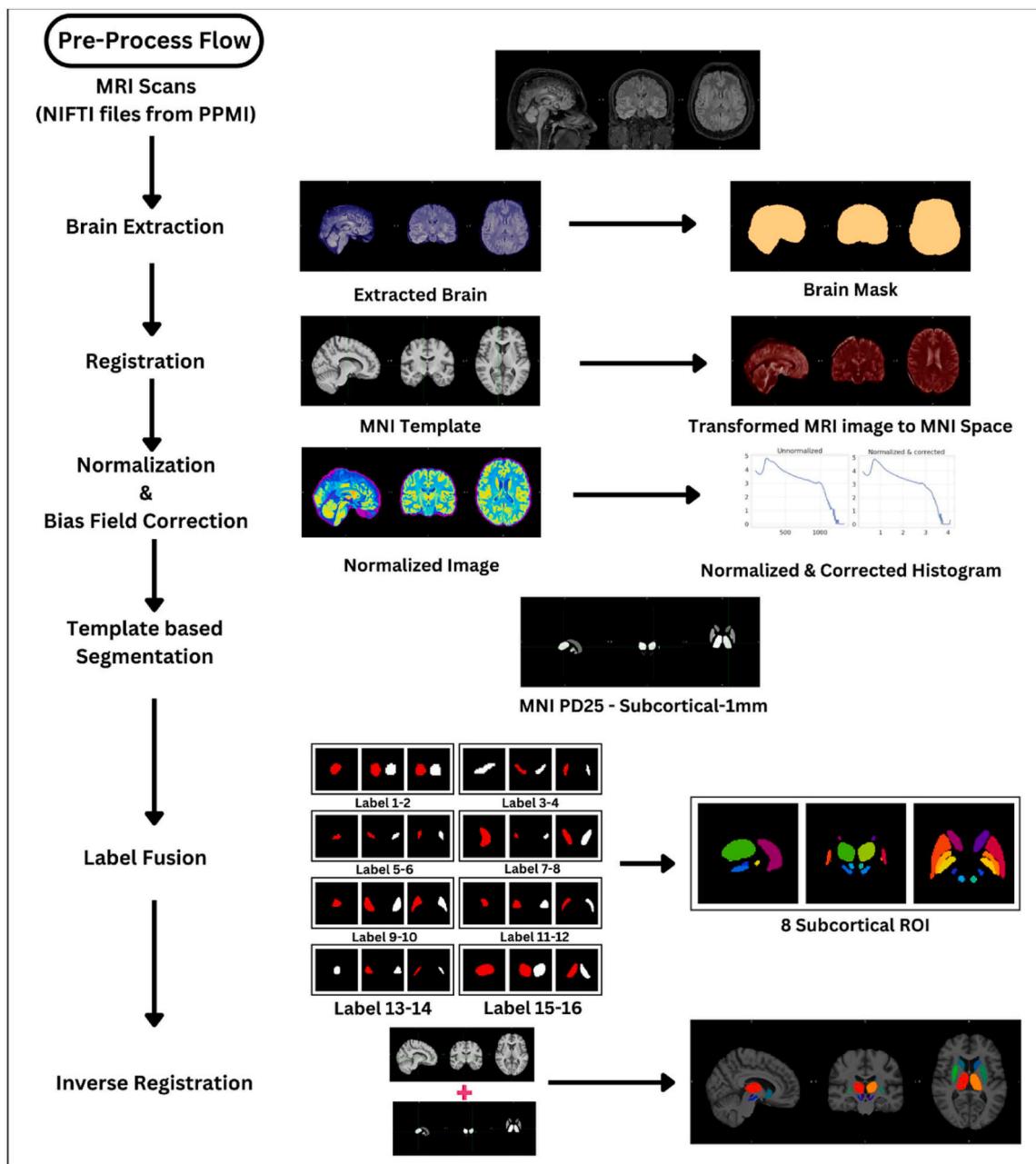


Fig. 2. Detail steps of Preprocessing.

3.2.2. Registration

The brain-extracted images are aligned to the MNIPD25-1T1MPRAGE-1 atlas using the Antspyx library, which ensures each image matches the atlas's spatial configuration. This library employs affine and non-linear registration techniques to align the images [38, 39]. The affine registration corrects for translations, rotations, and scaling, while the non-linear registration adjusts for local deformations. The normalized image from the registered image is obtained using Eqs. (1)-(3).

1. Minimum and maximum intensity value estimation.

$$\text{Min}_{\text{value}} = \min(I(x,y)); \text{Max}_{\text{value}} = \max(I(x,y)), \quad (1)$$

2. Subtraction of the minimum intensity value from each pixel.

$$\text{NormalizedImage}(x,y) = I(x,y) - \text{Min}_{\text{value}}, \quad (2)$$

3. Dividing each pixel in the normalized image by the range of intensity values.

$$\text{NormalizedImage}(x,y) = \frac{\text{NormalizedImage}(x,y)}{\text{Max}_{\text{value}} - \text{Min}_{\text{value}}} , \quad (3)$$

where I represent the T2 MRI image, $I(x,y)$ represents the image's intensity value at the pixel location (x,y) .

3.2.3. Bias Field Correction

The N4Bias correction algorithm is applied to each normalized image with the intended goal of correcting intensity inhomogeneities. This algorithm models and corrects low-frequency intensity non-uniformities by iteratively estimating the bias field and applying it to the image [40].

3.2.4. Subcortical segmentation

For each bias-corrected image, the FMRIB Software Library (FSL) tools are used to segment eight specific subcortical regions guided by the atlas. These tools utilize probabilistic atlases and machine learning techniques to segment regions accurately.

3.2.5. Label fusion

Used to improve segmentation accuracy, combining multiple segmentations via weighted voting based on the calculated probability of each label for each voxel, as represented in Eq. (4).

$$Y_i = \arg \max_{j \in L} \left(\sum_{k=1}^n S_{k,i,j} \cdot W_k \right) \quad (4)$$

where $S_{k,i,j}$ is the probability of voxel i belonging to label j in segmentation k , W_k is the weight assigned to segmentation, $\arg \max_j$ is the function that returns the label, L is the set of labels in the atlas, and Y_i is the final segmentation obtained by label fusion.

3.2.6. Inverse registration

The segmented regions are mapped back to the patient-specific anatomical space using inverse registration. This step involves using the transformation parameters obtained during the initial registration to align the segmented regions back to the original images. A detailed pseudocode and an algorithmic breakdown of the MRI preprocessing pipeline are described in Table 3. It includes the steps from converting DICOM files to performing advanced image segmentation using registration techniques.

3.3. Selection of subcortical regions

Dopamine-secreting neuron loss in the SN and Basal Ganglia (BG) is crucial for PD diagnosis. Distinct features of BG sub-regions also play a vital role in identifying PD. Affected PD patients typically show changes in cortical areas, altered sulcal patterns, and larger hippocampal volumes compared to controls. Neurotransmitters like glutamate, gamma-aminobutyric acid (found in the Subthalamic Nucleus), and acetylcholine (found in the Caudate) facilitate communication across brain areas,

Table 3

Pseudocode of MRI pre-processing pipeline.

Step 1: Initialize Processing

Input: NIfTI files for each patient

Step 2: Brain Extraction

For each NIfTI file:

 Apply the ROBEX algorithm to extract the brain tissue
 Generate a binary brain mask (1 for brain, 0 for non-brain)

Step 3: Registration to Atlas

For each brain-extracted image:

 Load the MNIPD25-T1MPRAGE-1 atlas
 Use Antspyx to align the image with the atlas

Step 4: Intensity Normalization

For each registered image:

 Calculate the intensity value using Eq. (1)
 Normalize the image using Eq. (2 and 3)

Step 5: Bias Field Correction

For each normalized image:

 Apply the N4Bias correction algorithm

Step 6: Subcortical Segmentation

For each bias-corrected image:

 Use FSL tools to segment eight subcortical regions based on the atlas

Step 7: Label Fusion

For each set of initial segmentations:

 For each voxel:
 Calculate the probability of each label using Eq. (4)

Step 8: Inverse Registration

For each segmented image:

 Apply inverse registration using parameters from Step 3
 Map segmented regions back to the original patient-specific anatomical space

End Algorithm

Output: Bias-corrected, registered, and segmented images

impacting cognition and motor functions. Specifically, the SN affects the caudate and putamen, which are vital for motor coordination, while the globus pallidus, thalamus, subthalamus, and red nucleus regulate motor and cognitive processes. These regions are critical for detecting PD and are used for further analysis [41–43].

3.4. Grading of Parkinson's disease

The grading process for PD presented in this study is illustrated in Fig. 3. This encompasses the use of atlas-based segmentation to identify 16 subcortical brain regions, allowing for a detailed analysis of their volume, morphology, and texture. Feature extraction methods are employed to identify critical patterns within the MRI data, which assist in differentiating between PD severity

classes [44]. The 'radiomics' package processes these patterns to refine them into a comprehensive set of features that represent the essential characteristics of the MRI data. The feature extraction process in Python using the pyradiomics package involves several steps. It begins with the loading of the pre-processed MRI image and its segmented label. Extraction settings are then configured, including voxel size resampling, normalization, and feature class selection. The pyradiomics extraction process is executed to compute the 107 features for each segmented region. This results in a comprehensive set of radiomic features critical for analyzing the progression of PD. The specific description of various types of radiomics features and the total number of features extracted from the subcortical regions, which is useful for the evaluation of the progression of PD, is presented in Table 4.

The feature selection selects the most significant features for the early detection of PD. In this study, a two-stage supervised feature selection process was utilized. Initially, Pearson's correlation method, with a threshold value of 0.85, was applied to the 107 radiomics features, effectively reducing the count to 39 by excluding 68 features that showed lesser relevance. Subsequently, the remaining 39 features were subjected to a wrapper-based Recursive Feature Elimination (RFE) procedure using an RF Classifier with 100 estimators and a random state of 42, which further refined the selection down to 20 features by excluding 19 features that did not significantly influence the prediction. Finally, a ranking model was applied to these 20 features to prioritize them based on their importance in predicting PD, as detailed in Fig. 4. This step ensured that the most impactful features were highlighted, equipping the ML model with the most discriminative data for accurate classification. To address the issue of imbalanced datasets in ML, this study applies a technique called SMOTE to ensure a balanced representation of all classes before model training [45]. This systematic approach, from detailed segmentation to balanced training, aims to improve the accuracy and dependability of PD classification.

3.5. Machine learning classifier and XAI techniques

This study focuses on selecting an optimal ML model for PD classification using the labelled dataset, emphasizing the need for XAI to provide accurate, robust, and understandable explanations of model predictions and behaviors. The dataset is used for both training and testing the ML model. As discussed in the previous section, the top 20 features are used as inputs for the ML models for PD classifications. In this research, six ML-based classifiers are assessed based on evaluation metrics, which include RF, KNN, Extra Trees (ET), Gradient boosting (GB), Decision Tree (DT), and CatBoost.

The study incorporates XAI techniques to explain the decisions made by the ML models. Among the XAI methods explored are Shapley additive explanations (SHAP), which provide a global measure of the feature importance. Local Interpretable Model-agnostic Explanations (LIME) and SHAPASH both offer detailed insights into the model's reasoning in arriving at a prediction. The weights of the surrogate model help to calculate the SHAP values that define the importance of each input feature. Eq. (5) is used to evaluate the SHAP value for individuals'

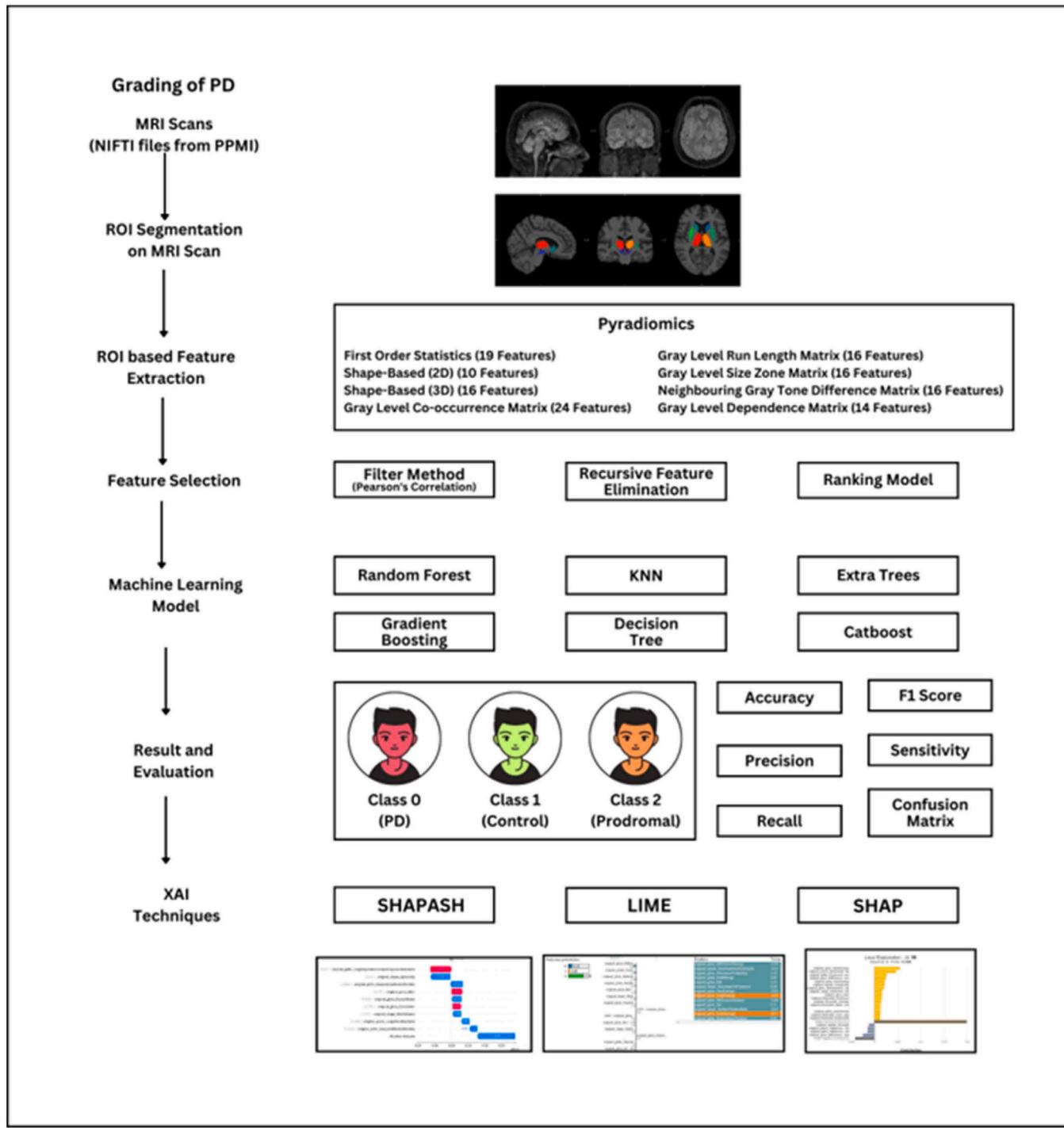


Fig. 3. Complete flow process proposed for Grading of PD.

instances.

$$\phi(x) = \sum_i \phi_i(x) \quad (5)$$

where $\phi(x)$ denotes the overall SHAP value for the instance x and $\phi_i(x)$ represents the SHAP value for a feature i in the instance x can be used to calculate the SHAP values that assign importance to each feature [46]. SHAP values are highly beneficial as they offer detailed insights into the model, giving important information on features that contribute to the prediction and making it transparent and valuable in clinical applications. Waterfall plots are employed for SHAP values, which show how

each feature contributes to the particular prediction, while beeswarm or bar plots display the feature's importance in general. Dependency plots explain the dependency of features on each other, and force plots help in understanding the prediction of a specific feature. Together, these visual tools aid in comprehending, troubleshooting, and refining the model.

LIME assesses prediction variations by introducing changes in input data. This methodology helps to understand the decision-making process of the black box model by offering insights into how these models behave locally [47]. Eq. (6) represents the mathematical expression of the LIME method for a specific instance.

$$\xi(x) = \operatorname{argmin}_{g \in G} L(f, g, w_x) + w(g) \quad (6)$$

Table 4

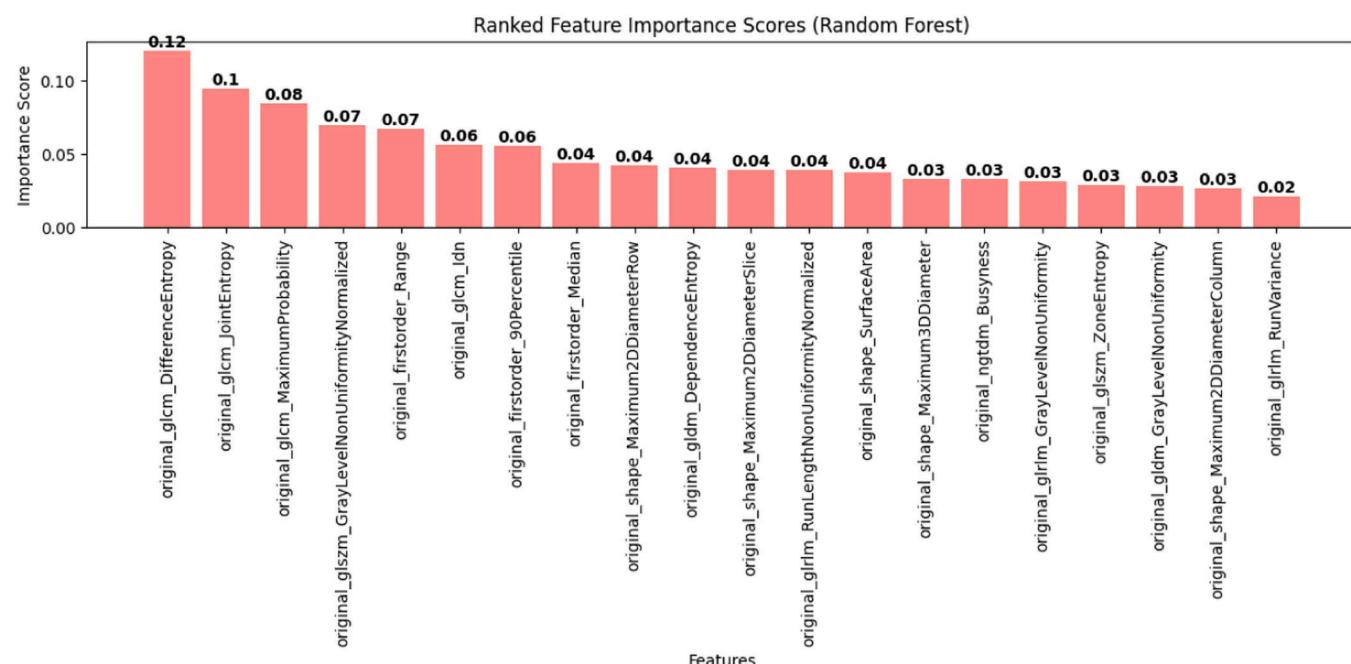
Explanation of features type and number of features extracted from ROI.

Feature Type	Number of Features	Explanation
First-order Statistical Features	17	It measures basic statistical properties of voxel intensity like mean, variance, and skewness, which are essential for detecting early PD-related changes in brain tissue.
Shape-Based (2D)	10	Analyses two-dimensional shape characteristics (area, perimeter, circularity) of subcortical structures; helpful in identifying morphological changes typical of PD.
Shape-based (3D)	10	It provides a detailed view of three-dimensional structural changes in the brain, which is crucial for diagnosing and assessing the progression of PD.
Gray Level Cooccurrence Matrix	24	Captures spatial relationships between voxel intensities, offering insights into textural changes in subcortical regions that are indicative of PD.
Gray Level Run Length Matrix	12	Describes the lengths of consecutive voxels with similar intensity values, helping to identify textural uniformity and variations associated with PD.
Gray Level Size Zone Matrix	12	Assesses the size of zones with uniform voxel intensities, which is helpful in detecting changes in tissue integrity and patterns associated with PD.
Neighbouring Gray Tone Difference Matrix	12	Measures intensity differences between neighboring voxels, highlighting textural and structural variations crucial for detecting subtle changes in the subcortical regions.
Gray Level Dependence Matrix	10	Analyzes dependencies among voxel intensities, revealing texture and structure details necessary for a comprehensive assessment of PD-related changes.
Exhaustive set of Features (Total)	107	A comprehensive set of features combining texture, shape, and structure analysis to characterize brain structures and detect PD robustly.

where $\xi(x)$ denotes the explainable model for an individual instance x , G represents the potential explanations, L denotes the loss function measuring the correspondence between the explanation model and the original model, f denotes the original model and w_x refers to the weighting factor assigned to each instance in the sampled data, which adjusts the influence of each instance based on its prevalence or importance in the original dataset. In cases where the sampled data closely resemble the original data, the importance w_x acquires greater values. SHAPASH delivers essential data and model information in addition to its visualization-focused approach. The variation in the model's predictions for various features, weighted by the Shapley values for the particular feature, are summed up in this method. These methods allowed us to create models that make precise predictions and provide easy-to-understand justifications for their choices.

4. Result

The comparative analysis in this study employed an ML classifier to investigate the textural, morphological, and statistical features of the Region of Interest extracted from MRI scans. Detailed information about the ML algorithms and their specific parameters is presented in Table 5. The software libraries used for the development of these models include CatBoost version 1.2.5 and Scikit-learn version 1.2.2. Radiomics features were extracted from a dataset of 500 subcortical segmented MR images. The data was split into a training set (80 %) and a testing set (20 %). The ML classifiers evaluated the performance metrics for detecting PD using true positives, true negatives, false positives, and false negatives. Eqs. (7) to (11) provide the mathematical formulas used to calculate these metrics, which help evaluate the effectiveness and accuracy of the models.

**Fig. 4.** Ranking the top 20 selected features.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{(\text{True Positives} + \text{False Positives} + \text{True Negatives} + \text{False Negatives})} \quad (7)$$

$$\text{Precision} = \frac{\text{True Positives}}{(\text{True Positives} + \text{False Positives})} \quad (8)$$

$$\text{Recall} = \frac{\text{True Positives}}{(\text{True Positives} + \text{False Negatives})} \quad (9)$$

$$\text{Specificity} = \frac{\text{True Negatives}}{(\text{False Positives} + \text{True Negatives})} \quad (10)$$

$$\text{F1score} = \frac{(2 * \text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (11)$$

After performing different ML models on the dataset, it was observed that the low performance was due to dataset imbalance. To address this, the SMOTE up-sampling method was employed to balance the classes and enhance model performance. [Table 6](#) illustrates the model's performance metrics before and after applying the up-sampling technique, demonstrating the effectiveness of SMOTE in improving data quality. Additionally, to showcase the distribution of features, a scatter plot is presented in [Fig. 5](#), which visually compares the feature distributions before and after using the up-sampling method.

This study evaluated the performance of various ML classifiers in detecting PD from brain MRI data, analyzing results across five separate test splits. The compared classifiers are RF, KNN, ET, GB, DT, and CatBoost, and their performances were measured using accuracy, precision, recall, specificity, F1 score, and computation time, as detailed in [Table 7](#). Among the classifiers, GB performed the best. It was configured with `n_estimators: 1000, learning_rate: 0.1, max_depth: 5, `min_samples_split: 2, `min_samples_leaf: 5` and loss: 'log_loss''. This configuration achieved the highest accuracy of 96.8 %, a precision of 97 %, a recall of 94.2 %, a specificity of 96.6 %, and an F1-score of 94.6 % in detecting PD. DT is the fastest, averaging minimal computation times of 33.0 seconds due to its straightforward node-splitting process. RF required longer times, averaging 84.9 seconds, due to the need to construct multiple trees. GB offered a balance between efficiency and accuracy, with average processing times of 76.78 seconds, making it the most effective classifier for detecting PD.

The performance of the GB classifier was evaluated through five-split cross-validation, with outcomes consistently surpassing other classifiers, as shown in [Fig. 6\(a\)](#). The confusion matrix for the GB model notably highlights its precision in identifying the prodromal stage of PD,

Table 5
Specifications used for different ML models.

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'

illustrating robust predictive accuracy for early-stage detection in [Fig. 6\(b\)](#). Receiver Operating Characteristics (ROC) curves for PD, control, and prodromal classes emphasize the model's exceptional discriminatory ability in [Fig. 6\(c\)](#). Specifically, the ROC curves show an Area Under the Curve (AUC) of 0.99 for both PD and control classes and a perfect AUC of 1.00 for the prodromal class, indicating flawless classification performance with a high True Positive Rate and a low False Positive Rate. GB model demonstrates high predictive accuracy, but a lack of explainability limits its clinical utility. For wider clinical adoption, it is essential to integrate XAI techniques that provide clarity on how decisions are derived, linking input features directly to predictive outcomes. This integration ensures that ML models are not only effective but also transparent and trustworthy, enhancing their practical utility in personalized patient care and promoting broader acceptance in the medical community.

To enhance transparency in AI-driven diagnostics, this study employs advanced XAI techniques, specifically SHAP and LIME. The Feature Importance Plot, as shown in [Fig. 7\(a\)](#), reveals the top features contributing to the model's predictions for class 2, with features such as Difference Entropy, Joint Entropy, and Difference Variance having the highest contributions. These features significantly impact the model's predictions, indicating areas for further investigation of their roles in the disease.

[Fig. 7\(b\)](#) demonstrates the interpretation of a prodromal case prediction using LIME, showing that the model predicts 18 % for PD, 8 % for Control, and 74 % for prodromal. The LIME explanation provides a local understanding of the prediction, with Difference Entropy and Joint Entropy features having significant positive contributions, while Surface Volume Ratio features have a negative contribution. This helps to validate the model's reasoning against expert knowledge by demonstrating how different features affect the prediction for the specific case. A global perspective of XAI is given by the SHAP summary plot in [Fig. 7\(c\)](#), which illustrates the influence of every feature across all predictions. Features like Difference Entropy and Joint Entropy consistently have high SHAP values, indicating their importance across many predictions. The plot also displays the contribution of each feature, with higher values of Difference Entropy leading to stronger impacts on predicting classes.

By integrating LIME and SHAP, this study provides both local and global explanations of the model's decision-making process. For example, LIME's local explanations showed that features such as the Difference Entropy had a significant influence/positive contribution to classifying the prodromal cases. SHAP's global explanations consistently highlighted the importance of key features, helping to identify which features are most influential across the entire dataset. These insights validate the model's predictions and also highlight the importance of difference entropy and joint entropy features in making appropriate decisions and in performing further clinical research. This also enhances

Table 6
Performance of the model before and after applying the up-sampling technique.

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 %

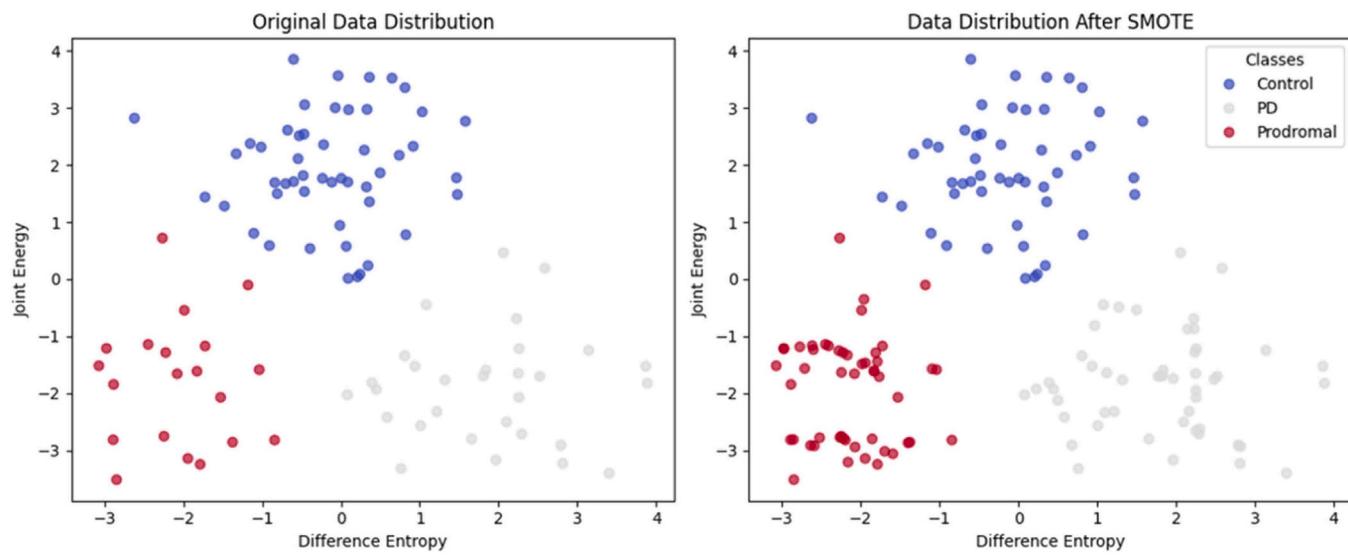


Fig. 5. Feature distribution plot using a scatter plot.

Table 7
Comparison of different machine learning classifiers.

Model	Split	Accuracy (in %)	Precision (in %)	Recall (in %)	Specificity (in %)	F1-score (in %)	Total 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 Boosting		95	96	92	95	94	78.63
Decision Trees		87	85	76	87	80	41.72
Catboost		90	94	84	93	88	43.86
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 Boosting		96	98	92	97	95	66.04
Decision Trees		89	90	89	88	90	30.84
Catboost		92	90	86	92	88	26.46
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 Boosting		98	98	98	97	95	83.51
Decision Trees		90	92	83	90	87	31.05
Catboost		91	92	83	91	87	33.29
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 Boosting		97	95	91	97	91	72.04
Decision Trees		92	84	93	88	88	30.68
Catboost		93	96	85	95	90	33.57
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 Boosting		98	98	98	97	98	83.72
Decision Trees		92	81	90	80	85	30.73
Catboost		94	98	87	97	92	33.45

transparency, trustworthiness, and actionable insights, particularly in sensitive domains like healthcare.

5. Discussion

This study analyzed neuroimaging features taken from brain scans for the early diagnosis of PD and its prodromal stages. Twenty significant features were selected, as shown in Table 8, and the selected features were statistically validated using ANOVA to assess differences across three groups: Control, PD, and Prodromal. The ANOVA results confirmed significant differences (p -values < 0.05) among all features, indicating their potential for distinguishing PD. Due to the inherent

biological variability and specialized nature of medical imaging data, these features might not follow a normal distribution and could exhibit varied variances among groups. Therefore, to ensure the robustness of our findings, we used the Kruskal-Wallis H-test after pairwise t-tests. This non-parametric test confirms that our findings are accurate. It is observed that even if the data isn't normally distributed, in order to avoid erroneous conclusions, pairwise t-tests and the Kruskal-Wallis H-test support our initial ANOVA results, showing that the selected features are effective in distinguishing between different stages of PD.

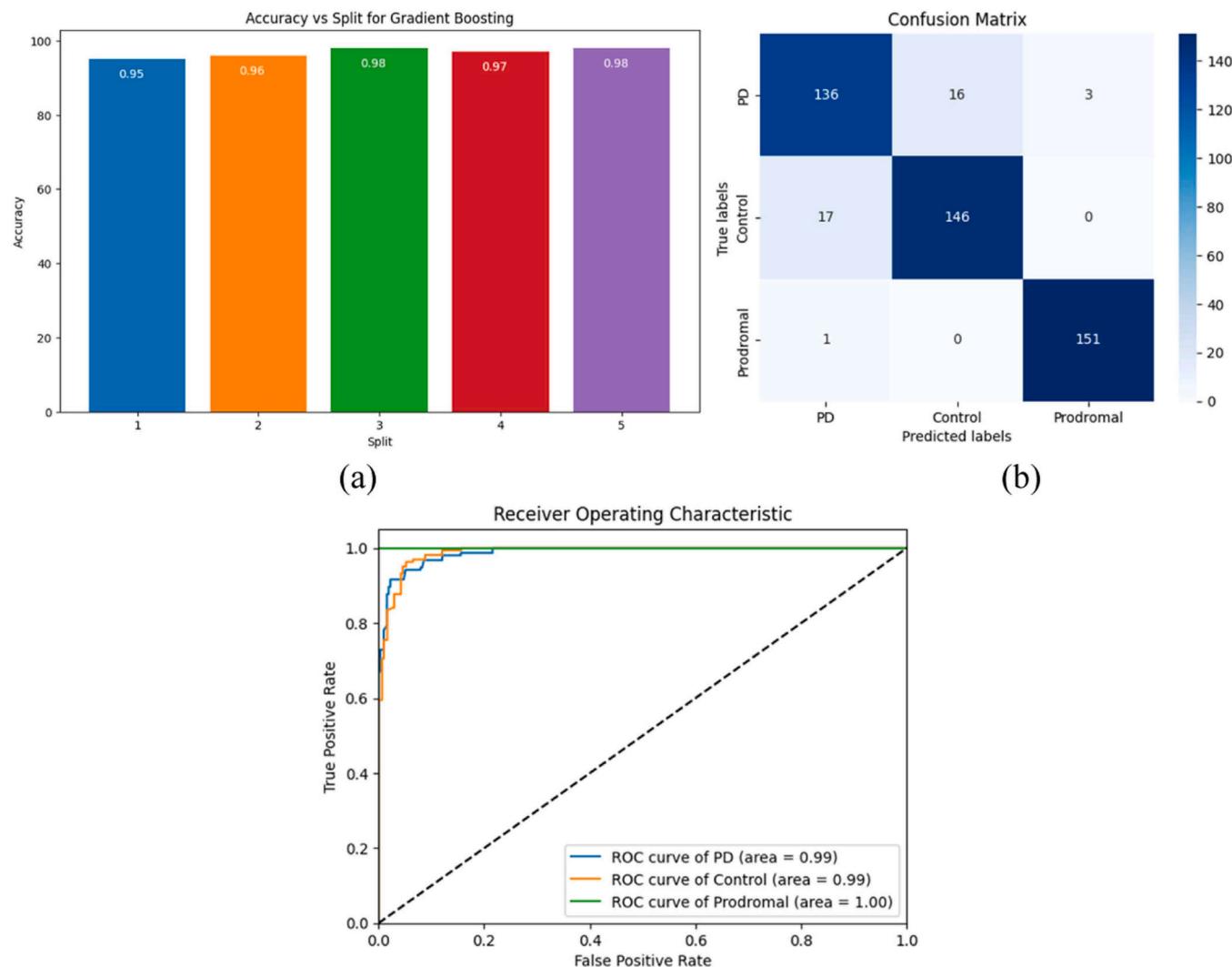


Fig. 6. (a) Accuracy of GB classifier across 5-Fold Cross-Validation splits, (b) Confusion matrix for GB classifier, (c) ROC curve three different classes for GB Classifiers model.

5.1. Significance of radiomics features in the diagnosis of PD

Detecting PD using radiomics features, such as measuring the Maximum Diameter to identify subcortical shrinkage, clearly acts as a biological marker. Similarly, reductions in Surface Area may suggest neuron loss due to degenerative changes. Intensity-based features that evaluate the brightness and range of brain images can indicate alterations from substances like iron, commonly found in PD-affected brains [48,49]. Features such as Gray Level Non-Uniformity and its Variance capture the uniformity of the brain's appearance; deviations of the features reflect the disease's impact on brain tissue, offering essential insights for early PD detection [50–52].

The physical interpretation of the selected imaging features provides significant insights into the progression of PD:

- a) Maximum 2D diameter column, Maximum 2D diameter row, Maximum 2D diameter slice, Maximum 3D diameter, and Surface: The numerical values of these shape features are smaller in the PD group, indicating shrinkage of brain structures due to the loss of neurons and other brain cells, resulting in reduced functionality and structural integrity of the subcortical regions. Intermediate values in the prodromal group suggest early structural changes. The control group has the most significant dimensions, with a much larger numerical value, reflecting healthy brain morphology.

b) First-order statistical features, including Median, 90th Percentile, and Range: These features describe the distribution of voxel intensities, with mean values of 1.1 for the control group, 0.4 for the prodromal group, and 0.1 for the PD group. Lower values in these features for the PD group indicate decreased intensity and potential neuronal loss in SN. Prodromal patients show intermediate values, hinting at early changes. The control group has higher values, indicating varied and healthy tissue intensities.

c) Texture analysis in the subcortical region: Features are extracted from the Gray Level Size Zone Matrix (GLSZM) and the Neighboring Gray Tone Difference Matrix (NGTDM). From NGTDM, key features such as Busyness, run variance, and grey level non-uniformity normalized are extracted, which indicates the variations in the texture in the subcortical region. In the PD group, lower values of entropy and higher values of maximum probability reflect less complex, more uniform textures, indicative of degeneration. The prodromal group shows intermediate values, suggesting the onset of changes. High-value entropy features and lower values in maximum probability features in the control group indicate complex, healthy tissue structures.

d) Texture Parameters Related to BG Size: Features such as Entropy in the Gray-Level Co-Occurrence Matrix (GLCM), Maximum Probability, and GLSZM feature indicate textural heterogeneity. Higher entropy indicates more significant textural heterogeneity and

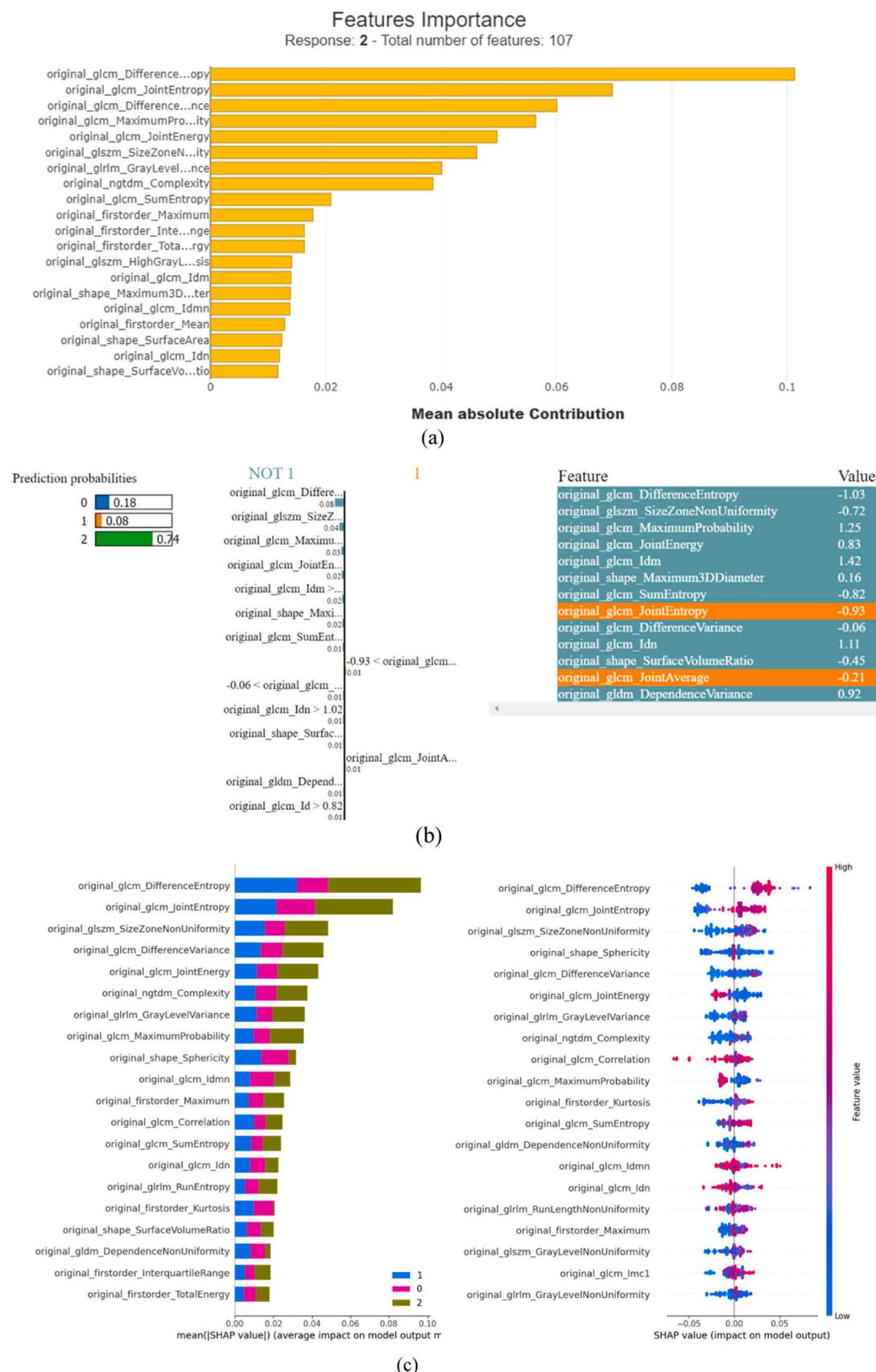


Fig. 7. Interpretation of the GB ML classifier for classification of PD disease (a) SHAPASH represent the local feature and its feature importance to individual instances (b) LIME shows the local importance of the feature by single instances (c) SHAP shows global and local feature importance and produce the values for every individual feature.

Table 8

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

complexity in the BG; maximum probability features indicate negative uniformity among adjacent pixels, indicating pathological changes. Higher GLSZM values denote less similarity and more textural heterogeneity.

Table 9 presents the inferences of the comparative analysis of the recent research in PD diagnosis using various ML models. This overview highlights significant improvements in the proposed method in comparison with the state-of-the-art techniques in terms of methodology and accuracy in the most recent times. This study evaluated various XAI methods, including SHAP, LIME, and SHAPASH, and determined that GB models provided the most reliable and clinically acceptable decisions. The incorporation of XAI methods highlighted the significance of the GLCM feature, specifically the “Difference Entropy” feature, as a key predictor of the GB classification. This approach enabled our models to produce accurate and interpretable results, which is crucial for practical medical applications.

5.2. Limitations and managerial implications

This study is oriented towards the deployment of XAI techniques and

their advantages in explaining the decisions made by ML models to provide reliable and trustworthy decisions for clinicians identifying the progression of PD. The developed framework provides robust predictions for PD classification using real-world data, aiding clinicians in better understanding the causes and availability of the disease. However, it has several limitations. Firstly, the developed framework uses the ANTs toolkit for MRI scan registration, which is a time-consuming process. This limitation can be addressed by incorporating more efficient methods like Voxel Morph [54], which can improve processing speed. Secondly, the investigation focuses on a specific ML model, providing a robust solution for real-world data. Incorporating models with automatic feature extraction capabilities, such as DL models, could identify more complex relationships in the features extracted from MRI scans, thus improving the prediction capability of the decision model.

Integrating the AI framework into existing healthcare information systems can assist neurologists and other clinicians in making informed decisions based on AI model outputs, leading to early and accurate diagnosis of PD. This could result in more timely and targeted interventions, significantly improving patients' quality of life. The implementation of advanced AI tools requires sophisticated technical and clinical skills. Therefore, medical organizations must plan for

Table 9
Comparison of PD Diagnostic Existing Methods with Our Proposed Approach.

Reference and Year	Dataset detail	Modality	Methodology	Region of Interest	Findings	Key observations
[9], 2020	203- HC 66- Prodromal 637-PD	MRI	ANN (Multilayer perceptron), Radiomics features extraction, 5-fold cross-validation	Subcortical regions	93.10 %- Accuracy, 95.41 %-Recall, 97.28 %- Precision, 94 %- F1-score	Advantage: High accuracy and precision Disadvantage: Lack of explainability
[29], 2023	36-HC 47-PD	MRI	SVM and RF, 10-fold cross-validation, radiomics features, custom model, 75 % of train, 25 % of test	SN region	83.33 %- accuracy	Advantage: Simple model, easier interpretation Disadvantage: Lower accuracy
[53], 2023	57-PD 74-MSA 70-HC	MRI and SWI scans	2640 radiomics features from both scans, LGBM difference between MSA and PD, as well as MSA, PD and HC, 5-fold cross-validation	subcortical regions	MSA vs PS vs HC- 81.4 %-accuracy, 88 %- AUC XAI techniques (i.e.) SHAP values used for feature importance	Advantage: Combining multiple modalities improves accuracy Disadvantage: More complex processing required
Proposed method	180-PD 160- Prodromal 160-HC	3D- MRI scans	GB + XAI (LIME, SHAP, SHAPASH) radiomics feature extraction, 2-fold feature selection method, SMOTE, statistical analysis, 5-fold cross-validation, 80 % of the train, 20 % of the test.	subcortical regions	96.8 % accuracy, 97 % precision, 94.2 % recall, 96.6 % specificity, 94.6 % F1-score	Advantage: High accuracy, comprehensive analysis, enhanced interpretability through XAI Disadvantage: Computationally intensive.

Abbreviations: SWI- Susceptibility Weighted Imaging, ANN- Artificial Neural Network, LGBM: Light Gradient Boosting Machine, MSA: Multiple System Atrophy

comprehensive training programs to ensure that medical staff can appropriately utilize and understand the AI system's decisions.

The proposed framework was developed and validated using real-world cases from the PPMI database, ensuring its applicability to actual clinical settings. The MRI scans and data used in developing this prediction strategy are from patients, guaranteeing that our results reflect practical, real-world scenarios. However, to further validate the performance of the proposed algorithms in real-world scenarios, the authors have taken initiatives through their affiliated university to establish collaborative tie-ups with leading hospitals. These efforts ensure that the findings of this work may be integrated into future studies and applications. Our framework accurately maps disease progression, but despite its reliability with XAI, it requires a more user-friendly interface for clinicians and patients. Early and precise diagnosis of PD is crucial, as the condition becomes challenging to manage at later stages, and precision in diagnosis is essential. Future work should focus on developing a suggestion model, such as a Large Language Model (LLM), to enhance diagnostic strategies and provide patient guidance. The model needs to be investigated further with more clinical validations to ensure its generalization characteristics. In this regard, the authors have initiated collaborations through their affiliated universities with hospitals to facilitate this process.

6. Conclusion and future directions

The proposed work developed an XAI-integrated ML classification model framework for diagnosing early PD. This framework not only improves disease prediction but also utilizes XAI techniques to provide insights into predictions and accurately identify the features responsible for those decisions, aiding in the preparation of effective diagnosis plans. Our study conducted a comparative analysis using six different ML methods on a real-world dataset of 500 images from PD, Control, and Prodromal classes. The proposed framework processes MRI scans using 3D slices to extract the radiomics features instead of 2D slices, avoiding the information loss inherent in traditional 2D methods. Among the extracted features, the top 20 that showed significant differences among the classes were identified using a 2-fold feature selection method. These top 20 features were validated by ANOVA (p -values < 0.05), demonstrating their effectiveness in distinguishing PD stages. The GB model achieved a superior accuracy of 96.8 % compared to other ML algorithms for the provided real-world data. XAI models were integrated into the GB model to elucidate the rationale behind its predictions. These XAI techniques highlighted crucial GLCM features (Difference entropy and Joint entropy), enhancing the model's interpretability and aiding clinicians in early PD diagnosis. GLCM features indicate brain tissue texture degradation and provide insights into the uniformity and structural changes in brain tissue, which is crucial for the early detection of PD. This study highlights the potential of XAI to enhance PD diagnosis and management, improving patient outcomes. It also emphasizes the need for targeted training programs to enable healthcare professionals to utilize AI insights effectively. To enhance the performance of the proposed PD diagnostic tools, future research can focus on exploring the feasibility of applying the multi-modal neuroimaging dataset and also deploying advanced AI technologies like DL and Reinforcement Learning. Future efforts can also incorporate more biomarkers and suitably modify XAI techniques to develop robust, clinically applicable solutions for personalized PD management and improved patient outcomes.

CRediT authorship contribution statement

S Priyadarshini: Writing – original draft, Methodology, Data curation, Conceptualization. **K Ramkumar:** Validation, Resources, Project administration, Formal analysis. **Subramaniyaswamy Vairavasundaram:** Writing – review & editing, Validation, Supervision, Software, Formal analysis. **K Narasimhan:** Writing – review & editing,

Validation, Formal analysis, Conceptualization. **S Venkatesh:** Validation, Methodology, Investigation, Data curation. **Rengarajan Amirtharajan:** Writing – review & editing, Validation, Methodology, Investigation. **Ketan Kotecha:** Supervision, Software, Resources, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data used in this study is available from the Parkinson's Progression Markers Initiative (PPMI), which is an open-access database.

Acknowledgments

The authors are grateful to the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), and Government of India for their financial support (CRG/2021/006697) and SASTRA Deemed University, India, for providing the necessary resources and support for the successful completion of this research. This work was supported by the Research Support Fund (RSF) of Symbiosis International (Deemed University), Pune, India.

References

- [1] W. Poewe, K. Seppi, C.M. Tanner, G.M. Halliday, P. Brundin, J. Volkmann, A.-E. Schrag, A.E. Lang, Parkinson disease, *Nat. Rev. Dis. Prim.* 3 (2017) 17013, <https://doi.org/10.1038/nrdp.2017.13>.
- [2] Z. Shu, P. Pang, X. Wu, S. Cui, Y. Xu, 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.* 12 (2020) 1–15, <https://doi.org/10.3389/fnagi.2020.548616>.
- [3] F. Salawu, A. Danburam, A. Olokoba, Non-motor symptoms of Parkinson's disease: diagnosis and management, *Niger. J. Med.* 19 (2010) 126–131, <https://doi.org/10.4314/njm.v19i2.56496>.
- [4] W. Zeng, J. Cai, L. Zhang, Q. Peng, Iron deposition in parkinson's disease: a mini-review, *Cell. Mol. Neurobiol.* 44 (2024), <https://doi.org/10.1007/s10571-024-01459-4>.
- [5] L. Ma, M. Gholam Azad, M. Dharmasivam, V. Richardson, R.J. Quinn, Y. Feng, D. L. Pountney, K.F. Tonissen, G.D. Mellick, I. Yanatori, D.R. Richardson, Parkinson's disease: alterations in iron and redox biology as a key to unlock therapeutic strategies, *Redox Biol.* 41 (2021), <https://doi.org/10.1016/j.redox.2021.101896>.
- [6] H.-M. Zeng, H.-B. Han, Q.-F. Zhang, H. Bai, Application of modern neuroimaging technology in the diagnosis and study of Alzheimer's disease, *Neural Regen. Res.* 16 (2021) 73, <https://doi.org/10.4103/1673-5374.286957>.
- [7] T. Mortezaeezadeh, H. Seyedarabi, B. Mahmoudian, J.P. Islamian, Imaging modalities in differential diagnosis of Parkinson's disease: opportunities and challenges, *Egypt. J. Radiol. Nucl. Med.* 52 (2021), <https://doi.org/10.1186/s43055-021-00454-9>.
- [8] N. Tambasco, F. Paolini Paoletti, A. Chiappinello, V. Lisetti, P. Nigro, P. Eusebi, P. Chiarini, M. Romoli, E. Brahimi, S. Simoni, M. Filidei, P. Floridi, R. Tarducci, L. Parnetti, P. Calabresi, T2*-weighted MRI values correlate with motor and cognitive dysfunction in Parkinson's disease, *Neurobiol. Aging* 80 (2019) 91–98, <https://doi.org/10.1016/j.neurobiolaging.2019.04.005>.
- [9] S. Chakraborty, S. Aich, 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* 8 (2020) 34, <https://doi.org/10.3390/healthcare8010034>.
- [10] M. Obayya, M.K. Saeed, M. Maashi, S.S. Alotaibi, A.S. Salama, M.A. Hamza, A novel automated Parkinson's disease identification approach using deep learning and EEG, *PeerJ Comput. Sci.* 9 (2023), <https://doi.org/10.7717/peerj-cs.1663>.
- [11] M.A. Islam, M.Z. Hasan Majumder, M.A. Hussein, K.M. Hossain, M.S. Miah, A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets, *Heliyon* 10 (2024) e25469, <https://doi.org/10.1016/j.heliyon.2024.e25469>.
- [12] S. Sivarajini, C.M. Sujatha, Morphological analysis of subcortical structures for assessment of cognitive dysfunction in Parkinson's disease using multi-atlas based segmentation, *Cogn. Neurodyn* 15 (2021) 835–845, <https://doi.org/10.1007/s11571-021-09671-4>.
- [13] B. Garcia Santa Cruz, A. Husch, 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.* 15 (2023), <https://doi.org/10.3389/fnagi.2023.1216163>.

- [14] A. Schrag, N. Quinn, How valid is the clinical diagnosis of Parkinson's disease, (2002) 529–534.
- [15] E. Tolosa, A. Garrido, S.W. Scholz, W. Poewe, M.D. Unit, N. Service, U. De Barcelona, Challenges in the diagnosis of Parkinson's disease, 20 (2022) 385–397. [https://doi.org/10.1016/S1474-4422\(21\)00030-2.Challenges](https://doi.org/10.1016/S1474-4422(21)00030-2.Challenges).
- [16] K.Y. Yamashita, S. Bhoopatiraju, B.D. Silvergate, G.T. Grossberg, Biomarkers in Parkinson's disease: a state of the art review, *Biomark. Neuropsychiatry* 9 (2023), <https://doi.org/10.1016/j.bionps.2023.100074>.
- [17] A. Govindu, S. Palwe, Early detection of Parkinson's disease using machine learning, *Procedia Comput. Sci.* 218 (2022) 249–261, <https://doi.org/10.1016/j.procs.2023.01.007>.
- [18] P.R. Magesh, R.D. Myloth, R.J. Tom, An Explainable Machine Learning Model for Early Detection of Parkinson's Disease using LIME on DaTSCAN Imagery, *Comput. Biol. Med.* 126 (2020) 104041, <https://doi.org/10.1016/j.combiomed.2020.104041>.
- [19] H.J. Yoon, K. Cho, W.G. Kim, Y.J. Jeong, J.E. Jeong, 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. (U.S.)* 100 (2021) E26961, <https://doi.org/10.1097/MD.00000000000026961>.
- [20] T. Pianpanit, S. Lolak, P. Sawangjai, T. Sudhawiyangkul, T. Wilaiprasitporn, Parkinson's disease recognition using SPECT image and interpretable AI: a tutorial, *IEEE Sens. J.* 21 (2021) 22304–22316, <https://doi.org/10.1109/JSEN.2021.3077949>.
- [21] D. Shi, X. Yao, Y. Li, H. Zhang, G. Wang, S. Wang, K. Ren, Classification of Parkinson's disease using a region-of-interest- and resting-state functional magnetic resonance imaging-based radiomics approach, *Brain Imaging Behav.* 16 (2022) 2150–2163, <https://doi.org/10.1007/s11682-022-00685-y>.
- [22] N. Islam, M.S.A. Turza, S.I. Fahim, R.M. Rahman, Advanced Parkinson's disease detection: a comprehensive artificial intelligence approach utilizing clinical assessment and neuroimaging samples, *Int. J. Cogn. Comput. Eng.* 5 (2024) 199–220, <https://doi.org/10.1016/j.ijcce.2024.05.001>.
- [23] H. Zhai, W. Fan, Y. Xiao, Z. Zhu, Y. Ding, C. He, W. Zhang, Y. Xu, Y. Zhang, Voxel-based morphometry of grey matter structures in Parkinson's Disease with wearing-off, *Brain Imaging Behav.* 17 (2023) 725–737, <https://doi.org/10.1007/s11682-023-00793-3>.
- [24] W. Yang, X. Bai, X. Guan, C. Zhou, T. Guo, J. Wu, X. Xu, M. Zhang, B. Zhang, J. Pu, J. Tian, The longitudinal volumetric and shape changes of subcortical nuclei in Parkinson's disease, *Sci. Rep.* 14 (2024) 1–9, <https://doi.org/10.1038/s41598-024-58187-4>.
- [25] M.R. Salmanpour, M. Shamsaei, A. Rahmim, Feature selection and machine learning methods for optimal identification and prediction of subtypes in Parkinson's disease, *Comput. Methods Prog. Biomed.* 206 (2021) 106131, <https://doi.org/10.1016/j.cmpb.2021.106131>.
- [26] D.A. Morales, Y. Vives-Gilabert, B. Gómez-Ansón, E. Bengoetxea, P. Larrañaga, C. Bielza, J. Pagonabarraga, J. Kulisevsky, I. Corcuera-Solano, M. Delfino, Predicting dementia classification in Parkinson's disease using Bayesian network classifiers, *Psychiatry Res. Neuroimaging* 213 (2013) 92–98, <https://doi.org/10.1016/j.psychresns.2012.06.001>.
- [27] H.-L. Chen, C.-C. Huang, X.-G. Yu, X. Xu, X. Sun, G. Wang, S.-J. Wang, An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach, *Expert Syst. Appl.* 40 (2013) 263–271, <https://doi.org/10.1016/j.eswa.2012.07.014>.
- [28] N. Betrouni, C. Moreau, A.S. Rolland, N. Carrière, M. Chupin, G. Kuchcinski, R. Lopes, R. Viard, L. Defebvre, D. Devos, Texture-based markers from structural imaging correlate with motor handicap in Parkinson's disease, *Sci. Rep.* 11 (2021) 1–10, <https://doi.org/10.1038/s41598-021-81209-4>.
- [29] S. Gore, A. Dhole, S. Kumbhar, J. Jagtap, A. Thakare, S. Gudadhe, Radiomics for Parkinson's disease classification using advanced texture-based biomarkers, *MethodsX* 11 (2023) 102359, <https://doi.org/10.1016/j.mex.2023.102359>.
- [30] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G.P.M. Van Stiphout, P. Granton, C.M.L. Zegers, R. Gillies, R. Boellard, A. Dekker, H.J.W.L. Aerts, Radiomics: Extracting more information from medical images using advanced feature analysis, *Eur. J. Cancer* 48 (2012) 441–446, <https://doi.org/10.1016/j.ejca.2011.11.036>.
- [31] P. Tupe-Waghmare, A. Rajan, S. Prasad, J. Saini, P.K. Pal, M. Ingalkhalikar, Radiomics on routine T1-weighted MRI can delineate Parkinson's disease from multiple system atrophy and progressive supranuclear palsy, *Eur. Radiol.* 31 (2021) 8218–8227, <https://doi.org/10.1007/s00330-021-07979-7>.
- [32] Z. Yu, K. Wang, Z. Wan, S. Xie, Z. Lv, Popular deep learning algorithms for disease prediction: a review, *Clust. Comput.* 26 (2023) 1231–1251, <https://doi.org/10.1007/s10586-022-03707-y>.
- [33] M. Camacho, M. Wilms, P. Mouches, H. Almgren, R. Souza, R. Camicioli, Z. Ismail, O. Monchi, N.D. Forkert, Explainable classification of Parkinson's disease using deep learning trained on a large multi-center database of T1-weighted MRI datasets, *NeuroImage Clin.* 38 (2023) 103405, <https://doi.org/10.1016/j.nicl.2023.103405>.
- [34] M. Nazari, A. Kluge, I. Apostolova, S. Klutmann, S. Kimiae, M. Schroeder, R. Buchert, 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* 49 (2022) 1176–1186, <https://doi.org/10.1007/s00259-021-05569-9>.
- [35] J.J. Bouza, C.-H. Yang, D. Vaillancourt, B.C. Vemuri, MVC-Net: A Convolutional Neural Network Architecture for Manifold-Valued Images With Applications, (2020). <http://arxiv.org/abs/2003.01234>.
- [36] X. Li, P.S. Morgan, J. Ashburner, J. Smith, C. Rorden, The first step for neuroimaging data analysis: DICOM to NIfTI conversion, *J. Neurosci. Methods* 264 (2016) 47–56, <https://doi.org/10.1016/j.jneumeth.2016.03.001>.
- [37] S.M. Smith, Fast robust automated brain extraction, *Hum. Brain Mapp.* 17 (2002) 143–155, <https://doi.org/10.1002/hbm.10062>.
- [38] Y. Xiao, V. Fonov, M.M. Chakravarty, S. Beriault, F. Al Subaie, A. Sadikot, G. B. Pike, G. Bertrand, D.L. Collins, A dataset of multi-contrast population-averaged brain MRI atlases of a Parkinson's disease cohort, *Data Br.* 12 (2017) 370–379, <https://doi.org/10.1016/j.dib.2017.04.013>.
- [39] L.J. Isaksson, S. Raimondi, F. Botta, M. Pepa, S.G. Gugliandolo, S.P. De Angelis, G. Marvaso, G. Petralia, O. De Cobelli, S. Gandini, M. Cremonesi, F. Cattani, P. Summers, B.A. Jereczek-Fossa, Effects of MRI image normalization techniques in prostate cancer radiomics, *Phys. Med.* 71 (2020) 7–13, <https://doi.org/10.1016/j.ejmp.2020.02.007>.
- [40] N.J. Tustison, B.B. Avants, P.A. Cook, Yuanjie Zheng, A. Egan, P.A. Yushkevich, J. C. Gee, N4ITK: improved N3 bias correction, *IEEE Trans. Med. Imaging* 29 (2010) 1310–1320, <https://doi.org/10.1109/TMI.2010.2046908>.
- [41] J. Li, Y. Zhang, Z. Huang, Y. Jiang, Z. Ren, D. Liu, J. Zhang, R. La Piana, Y. Chen, Cortical and subcortical morphological alterations in motor subtypes of Parkinson's disease, *Npj Park. Dis.* 8 (2022) 167, <https://doi.org/10.1038/s41531-022-00435-3>.
- [42] C.J. Park, J. Eom, K.S. Park, Y.W. Park, S.J. Chung, Y.J. Kim, S.S. Ahn, J. Kim, P. H. Lee, Y.H. Sohn, S.-K. Lee, An interpretable multiparametric radiomics model of basal ganglia to predict dementia conversion in Parkinson's disease, *Npj Park. Dis.* 9 (2023) 127, <https://doi.org/10.1038/s41531-023-00566-1>.
- [43] J.-H. Deng, H.-W. Zhang, X.-L. Liu, H.-Z. Deng, 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* 12 (2022) 1356–1366, <https://doi.org/10.5498/wjp.v12.i12.1356>.
- [44] Y. Wu, J.-H. Jiang, L. Chen, J.-Y. Lu, J.-J. Ge, F.-T. Liu, J.-T. Yu, W. Lin, C.-T. Zuo, J. Wang, Use of radiomic features and support vector machine to distinguish Parkinson's disease cases from normal controls, *Ann. Transl. Med.* 7 (2019), <https://doi.org/10.21037/atm.2019.11.26>.
- [45] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, *J. Artif. Intell. Res.* 16 (2002) 321–357, <https://doi.org/10.1613/jair.953>.
- [46] S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, *Adv. Neural Inf. Process. Syst.* 2017–Decem (2017) 4766–4775, <http://arxiv.org/abs/1705.07874>.
- [47] M. Ribeiro, S. Singh, C. Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in: *Proc. 2016 Conf. North Am. Chapter Assoc. Comput. Linguist. Demonstr.*, Association for Computational Linguistics, Stroudsburg, PA, USA, 2016: pp. 97–101. <https://doi.org/10.18653/v1/N16-3020>.
- [48] S.J. Chung, H.S. Lee, H.S. Yoo, Y.H. Lee, P.H. Lee, Y.H. Sohn, Patterns of striatal dopamine depletion in early Parkinson disease: prognostic relevance, *Neurology* 95 (2020) E280–E290, <https://doi.org/10.1212/WNL.0000000000009878>.
- [49] L.I. Wallis, M.N.J. Paley, J.M. Graham, R.A. Grünewald, E.L. Wigñall, H.M. Joy, P. D. Griffiths, MRI assessment of basal ganglia iron deposition in Parkinson's disease, *J. Magn. Reson. Imaging* 28 (2008) 1061–1067, <https://doi.org/10.1002/jmri.21563>.
- [50] D. Molina, J. Pérez-Beteta, A. Martínez-González, J. Martíno, C. Velásquez, E. Arana, V.M. Pérez-García, Influence of gray level and space discretization on brain tumor heterogeneity measures obtained from magnetic resonance images, *Comput. Biol. Med.* 78 (2016) 49–57, <https://doi.org/10.1016/j.combiomed.2016.09.011>.
- [51] N.Y. Shin, M. Bang, S.W. Yoo, J.S. Kim, E. Yun, U. Yoon, K. Han, K.J. Ahn, S.K. Lee, Cortical thickness from MRI to predict conversion from mild cognitive impairment to dementia in Parkinson disease: a machine learning-based model, *Radiology* 300 (2021) 390, <https://doi.org/10.1148/radiol.2021203383>.
- [52] N.J. Tustison, J. Gee, Run-length matrices for texture analysis, *Insight J.* 1 (2022) 1–6, <https://doi.org/10.54294/ex0itu>.
- [53] S. Bu, H. Pang, X. Li, M. Zhao, J. Wang, Y. Liu, H. Yu, Multi-parametric radiomics of conventional T1 weighted and susceptibility-weighted imaging for differential diagnosis of idiopathic Parkinson's disease and multiple system atrophy, *BMC Med. Imaging* 23 (2023) 1–10, <https://doi.org/10.1186/s12880-023-01169-1>.
- [54] G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, VoxelMorph: a learning framework for deformable medical image registration, *IEEE Trans. Med. Imaging* 38 (2019) 1788–1800, <https://doi.org/10.1109/TMI.2019.2897538>.