

Article

Bridging Accuracy and Interpretability: Explainable Deep Learning for Parkinson's Disease Diagnosis from MRI

Ioana-Teodora Isar and Nirvana Popescu * 

Computer Science and Engineering Department, National University of Science and Technology POLITEHNICA Bucharest, 060042 Bucharest, Romania; ioana_teodora.isar@stud.acs.upb.ro

* Correspondence: nirvana.popescu@upb.ro

Abstract

This paper presents a deep learning framework for automated Parkinson's disease (PD) detection from T1-weighted MRI scans using the NTUA dataset. The proposed pipeline combines advanced preprocessing, robustness evaluation, and explainable AI techniques to improve both diagnostic performance and clinical interpretability. Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to enhance anatomical details, while SMOTE was applied to the deep feature vectors extracted from the training MRI images to address class imbalance. Several state-of-the-art convolutional neural networks were evaluated through six patient-wise train–test splits to assess robustness and generalization across subjects. In addition, a targeted experiment using only axial MRI slices from 50 subjects was conducted to reduce irrelevant anatomical information and emphasize brain regions potentially associated with PD. Among the evaluated models, performance varied substantially across subject-wise splits, highlighting a strong dependency on patient selection. While peak configurations reached high individual metrics, the aggregate subject-wise analysis demonstrated a more conservative baseline. To improve transparency, Grad-CAM visualizations were generated, showing that the models primarily focused on central brain structures relevant to Parkinsonian neurodegeneration, with minimal attention extending to non-diagnostic regions. The results demonstrate that combining adaptive preprocessing, patient-wise evaluation, and explainable deep learning holds promise for MRI-based Parkinson's disease detection under a preliminary, dataset-specific evaluation, though substantial performance variability remains across different subject selections, rather than demonstrating immediate clinical readiness.

Keywords: Parkinson's disease; MRI; deep learning; Grad-CAM



Academic Editors: Kalpdrum Passi and Francesc Pozo

Received: 23 June 2026

Revised: 7 August 2026

Accepted: 9 August 2026

Published: 13 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Parkinson's Disease and Early Diagnosis

Parkinson's disease affects approximately 2–3% of the population, particularly elderly individuals, and represents the second most common neurodegenerative disorder after Alzheimer's disease. It is primarily characterized by motor impairments such as bradykinesia (slowness of movement), resting tremor, muscular rigidity, and postural instability, which progressively reduce patients' mobility and quality of life [1]. In addition to these motor manifestations, Parkinson's disease is frequently associated with a wide spectrum of non-motor symptoms, including cognitive decline, autonomic dysfunction, sleep disturbances, depression, anxiety, and sensory abnormalities. These symptoms often appear in

the early stages of the disease and may significantly impact daily functioning even before clear motor deficits become clinically evident.

The progressive nature of Parkinson's disease and the absence of definitive laboratory biomarkers make early diagnosis particularly challenging. Current clinical assessment methods rely heavily on neurological examination and symptom observation, which can lead to delayed or subjective diagnosis, especially during the prodromal stages of the disease. Consequently, there is an increasing interest in the development of computer-aided diagnostic systems capable of supporting clinicians through objective and data-driven analysis.

Recent advances in medical imaging and artificial intelligence have opened new perspectives for the automated detection of neurodegenerative disorders. In particular, deep learning methods applied to magnetic resonance imaging (MRI) data have demonstrated promising capabilities in identifying subtle structural brain alterations associated with Parkinson's disease. These approaches have the potential to improve diagnostic accuracy, enable earlier intervention, and support more personalized treatment strategies, ultimately contributing to better patient management and long-term clinical outcomes.

1.2. Deep Learning for Parkinson's Disease Identification

Machine learning has become an effective method for detecting this disease from the early stages using several types of data collected from patients associated with various symptoms. These are divided into voice analysis, sensor data, medical images and biometric data.

In terms of changes occurring in the brain, Parkinson's disease is characterized by the degeneration of dopaminergic neurons, thus the reduction in dopamine in the striatum, hippocampal atrophy associated with memory impairment, and atrophy of the prefrontal cortex, affecting executive functions [2].

The main challenges in using machine learning are the small size of the collected datasets, which leads to overfitting, and the choice of imaging methods, since techniques such as PET (Positron Emission Tomography) and SPECT (Single-Photon Emission Computed Tomography) are invasive for the patient, who is subjected to a level of radiation. For these reasons, structural magnetic resonance imaging is used in the diagnosis of the disease because it highlights the contrast between different types of tissue, making it possible to detect structural changes in the substantia nigra [3].

1.3. Contributions

This paper presents a clinically oriented deep learning framework for automated Parkinson's disease (PD) detection from T1-weighted MRI scans, integrating advanced preprocessing, explainability, and robustness evaluation strategies. Using the NTUA neuroimaging dataset, we investigate the effectiveness of multiple state-of-the-art convolutional neural network architectures for PD classification and introduce a comprehensive evaluation pipeline designed to improve both predictive performance and clinical interpretability.

To enhance discriminative anatomical features, Contrast Limited Adaptive Histogram Equalization (CLAHE) was employed for adaptive contrast enhancement, while Synthetic Minority Oversampling Technique (SMOTE) was applied to the deep feature vectors extracted from the training partition in order to balance the minority class. Beyond standard classification experiments, we propose a targeted axial-slice analysis strategy that isolates brain regions potentially associated with Parkinson's disease and reduces the influence of anatomically irrelevant information. This focused representation enabled more efficient learning from only 50 subjects while preserving strong diagnostic performance.

Extensive comparative experiments were conducted using six patient-wise train–test configurations to assess model robustness, inter-subject variability, and generalization capability—an aspect often overlooked in previous studies. Among the evaluated architectures, ResNet50 achieved the highest overall classification accuracy of 95%, whereas DenseNet121 demonstrated superior clinical sensitivity by attaining a 99% recall for the Parkinson’s disease class, highlighting its potential suitability for screening-oriented applications where minimizing false negatives is critical.

To further improve transparency and trustworthiness of the proposed framework, Grad-CAM explainability analysis was incorporated, revealing that the networks consistently focused on central brain structures associated with Parkinsonian neurodegeneration. These visualizations provide qualitative insights

into the anatomical regions influencing the decision-making process and support the interpretability of deep learning predictions in neuroimaging applications.

The obtained results demonstrate that combining adaptive preprocessing, patient-wise robustness analysis, and explainable AI techniques can substantially improve MRI-based Parkinson’s disease detection. Also, a separate experiment based on multiple subject-wise data splits was conducted to assess the robustness of the models and the influence of patient selection on the reported performance. The proposed framework establishes a foundation for more reliable and clinically interpretable computer-aided diagnostic systems, while future work will focus on validation across larger multi-center datasets to further strengthen model generalization and translational applicability.

2. Related Work on MRI-Based Parkinson’s Disease Identification

Over the years, multiple deep learning approaches have been proposed for the automatic detection of Parkinson’s disease from structural MRI data. Recent literature underscores the need for robust neuroimaging preprocessing. The review [4] by Noor et al. shows that structural MRIs are prone to noise and intensity variations, and the authors support contrast enhancement tools like CLAHE. Additionally, the study addresses how imbalanced datasets degrade deep learning performance, suggesting oversampling to mitigate class bias. In the section dedicated to Parkinson’s disease detection, the authors validate CNN backbones like ResNet for neurodegenerative disease classification. Another study [5], evaluated the performance of three CNN architectures (InceptionV3, VGG16, and VGG19) for Parkinson’s disease detection from MRI slices and utilized the Parkinson’s Progression Markers Initiative (PPMI) dataset. The comparative analysis demonstrated that VGG19 achieved the highest diagnostic accuracy at 91.5%, outperforming InceptionV3 (89.5%) and VGG16 (88.5%). Similarly, the study [6] utilized a pretrained AlexNet architecture to classify Parkinson’s disease from MRI scans. The authors expanded their dataset up to 4200 images using GAN-based augmentation and adopted a 60/20/20 data split for training, validation, and testing. Their transfer learning approach achieved an average accuracy of 89.23%.

NTUA Parkinson’s Dataset was created in the absence of standardized datasets by the NTUA team [7]. The images from the MRI Axial T1 sequence and Dat Scans were used in the study. The proposed model consists of an end-to-end architecture with ResNet50 (CNN), for extracting important features from the images, and RNN (LSTM/GRU), for obtaining predictions. The maximum accuracy, 98%, was obtained by combining the two types of images, MRI and Dat Scans with a recall of 96% and precision of 100%.

The paper [8] proposed an architecture based on ResNet50 with Channel-Specific Attention Module (CSAM) and used the NTUA and PPMI (Parkinson’s Progression Markers Initiative) datasets. The training dataset contains 80% of the initial one and the testing 20%.

The preprocessing methods consist of resizing and normalization and the CSAM module was added at the architectural level.

The CSAM module is inserted between the convolutional and batch normalization layers, to identify important areas in the image. The Channel attention part analyzes the channels extracted (edges, textures, shapes) from the image by the previous layers, assigns values according to relevance and builds a feature map. In the second stage, the module combines the information from the channels and identifies important areas.

The proposed model achieved an accuracy of 99.16%, comparable to the best result reported in [8], where ResNet50 combined with a CSAM attention module also reached 99.16%. In contrast, DenseNet121 and MobileNetV3 achieved lower accuracies of 90.7% and 96.53%, respectively, while the optimized MobileNetV3 obtained 98.36%.

The article [9] defined a convolutional neural network PDCNet that includes: 3 convolutional layers, pooling layers, dropout with different rates, 2 dense layers and a softmax layer for the output. The authors applied Grad-CAM to be able to interpret the model predictions. The performance comparison of the PDCNet architecture for different dropout rates is presented in Table 1.

Table 1. Performance comparison of the PDCNet architecture on the NTUA dataset for different dropout rates [9].

Dropout	Recall (%)	Specificity (%)	Accuracy (%)
0.1	95.42	97.03	96.23
0.2	96.85	98.27	97.56
0.3	97.69	98.85	98.27
0.4	97.21	98.49	97.85

On the other hand, a different approach is the work of Chakraborty et al. in the study [10]. They trained a specialized 3D CNN using MRI scans from 406 subjects in the PPMI database. Their model hit an accuracy of 95.29% and a recall of 94.21%. Their visual explanation maps confirmed that the model attended to the exact biological markers of Parkinson's disease, specifically neuron loss in the substantia nigra. However, they suggested focusing future efforts on specific subcortical regions, which makes a strong case for investigating efficient 2D models that evaluate slices at the patient level.

3. Materials and Methods

3.1. Dataset

The study uses T1-weighted MRI scans from the standardized NTUA Parkinson's dataset, proposed by the NTUA team [7]. It contains 43,087 MRI images and 920 DaT Scans images obtained from 78 subjects (55 patients diagnosed with Parkinson's and 23 healthy subjects). For each subject, multiple sections obtained through MRI examinations and DaT Scans are available.

Before the experiments, all MRI images were manually inspected to ensure that only images of adequate quality containing the anatomical regions relevant to Parkinson's disease analysis were included. MRI slices presenting severe motion artifacts, strong intensity corruption, incomplete brain coverage, or anatomical sections without visible midbrain structures were excluded. In addition, slices containing predominantly superior or inferior brain regions that did not include the substantia nigra were discarded. After applying these quality criteria, 67 subjects remained in the dataset. For the subject-wise robustness experiment, only axial slices covering the midbrain region were retained for each subject, resulting in a subset of 50 subjects. The midbrain region was selected because it includes the substantia nigra, which is strongly associated with Parkinson's disease pathology.

The preprocessing stage consisted of dividing the dataset according to the MRI technique used to keep only T1 sequence images. MRI scans were standardized using MONAI (version 1.5.2) through foreground cropping, intensity normalization, and resizing to a fixed resolution of 224×224 pixels.

The train/test split was performed at the subject level using a fixed random seed of 42, with approximately 90% of the subjects assigned to the training partition and 10% to the independent testing partition. To prevent data leakage, all MRI slices belonging to the same subject were assigned exclusively to a single partition. The validation partition was subsequently generated from the training partition by randomly selecting 20% of the training subjects using a fixed random seed of 4. After this stage, as depicted in Table 2, the training partition contains 50 subjects (7159 slices), the validation partition includes 11 subjects (1343 slices), and the independent testing partition consists of 6 subjects (810 slices).

Table 2. Distribution of subjects and MRI slices across the training, validation, and testing partitions.

Partition	Class Label	No. of Subjects	No. of Slices (Images)
Training	PD Patients	35	5190
	Non-PD Patients	15	1969
Validation	PD Patients	8	902
	Non-PD Patients	3	441
Testing	PD Patients	4	355
	Non-PD Patients	2	455
Total	–	67	9312

Training images were augmented using ImageDataGenerator with `rotation_range = 20` degrees, `width_shift_range = 0.2`, `height_shift_range = 0.2`, `shear_range = 0.2`, `zoom_range = 0.2`, `horizontal_flip = True`, and `fill_mode = 'nearest'`. No augmentation was applied to the validation or test partitions. All images were resized to 224×224 pixels using bilinear interpolation, with a batch size of 32 and a fixed generator seed of 42. Training data were shuffled, whereas validation and test data were not shuffled.

3.2. Image Enhancement

To increase the contrast of the image, the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique was used, which divides the received image into sections and applies histogram equalization to each portion [11].

The parameters were varied to see which case obtains the best results. The first parameter, `clipLimit`, indicates the limit of contrast changes and `tileGridSize` sets the number of image divisions. The process consists of applying CLAHE to the grayscale and normalized image. The `clipLimit` parameter was evaluated within the range $[0.01, 1.0]$, while `tileGridSize` values of $\{16, 32\}$ were tested. The resulting image was assigned a score based on sharpness and noise. The best-performing configuration was obtained with `clipLimit = 0.1` and `tileGridSize = 16`.

3.3. Data Balancing

The Parkinson's medical imaging dataset is obtained from a small sample of people, most of whom suffer from the disease. Thus, the deep learning models are affected by data imbalance and overfitting may occur. In this study, the balancing of the dataset was performed using the Synthetic Minority Oversampling Technique (SMOTE). The technique was applied exclusively to the training partition after the validation and testing sets had been separated. A fixed random seed of 42 was used for the procedure to ensure

reproducibility. The MRI images were first processed by the pretrained convolutional neural network to extract deep feature vectors. These feature vectors were standardized using StandardScaler, after which SMOTE generated synthetic feature vectors for the minority class by interpolating between neighboring samples in the feature space [12]. The balanced feature vectors were subsequently used to train the classifier, while preserving the original classification architecture.

3.4. Models

The proposed comparison uses four convolutional neural network architectures that have achieved positive results in the field of medical imaging and are available for transfer learning: DenseNet121 [13], ResNet50 [14], ResNet101 [14], and VGG16 [15]. A framework of the proposed system is shown in Figure 1.

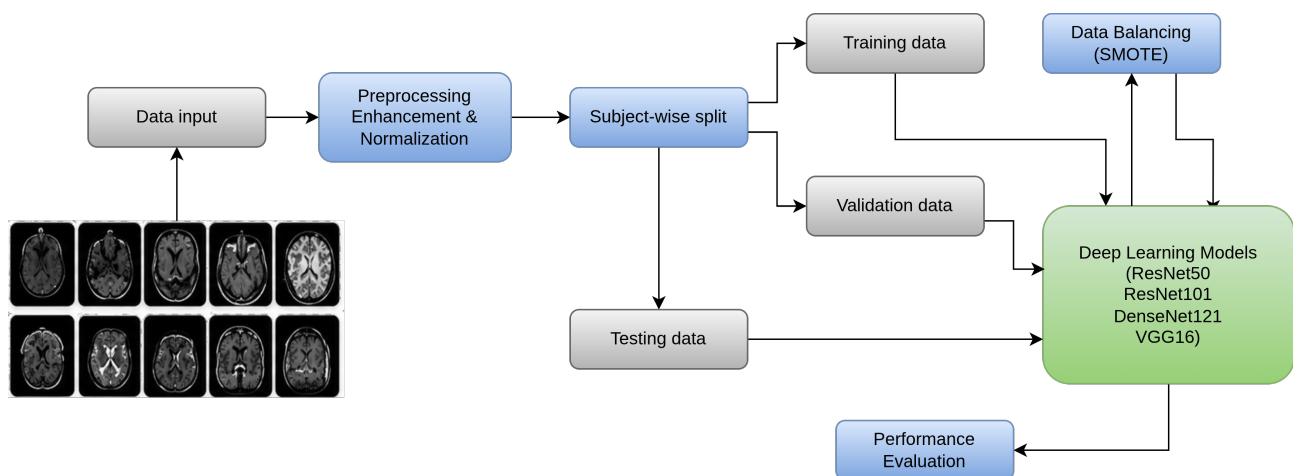


Figure 1. Framework of the proposed system.

All CNN backbones were initialized with ImageNet-pretrained weights and adapted to the binary MRI classification task. The final fully connected layer of each network was replaced with a dense layer containing a Softmax activation to enable binary classification.

The first convolutional layers mainly capture generic image features relevant to any type of MRI image, such as edges, textures, and intensity patterns. Therefore, only the deeper layers were fine-tuned to learn Parkinson’s disease-specific features while reducing the risk of overfitting on a relatively small dataset. For DenseNet121, the last 100 layers were unfrozen, whereas for ResNet50 and ResNet101, the first 140 and 250 layers were frozen, respectively. For VGG16, only the first 5 layers were kept frozen during training.

The selected fine-tuning strategies were determined experimentally. For DenseNet121, preliminary experiments showed that training all layers increased the risk of overfitting, whereas unfreezing only the last 30 layers limited the model’s ability to adapt. Unfreezing the last 100 layers provided a good compromise between preserving the representations learned from ImageNet and adapting the network to Parkinson’s disease MRI classification.

VGG16 has a simplified architecture compared to the other models. For this reason, two dense layers containing 512 and 128 neurons were added in order not to lose any of the features obtained and the dropout rates of 0.2 and 0.3 were used.

In the case of ResNet50, the last layers have considerable dimensions and in order not to drastically reduce the information retained, two dense layers with 256 and 128 neurons were added, but the dropout was increased to 0.4 and 0.3 compared to VGG16.

The architecture of DenseNet121 allowed the addition of a single dense layer with 128 neurons and a dropout of 0.5 at the end. The motivation is the increased complexity of the model and the avoidance of overfitting in the case of a limited dataset.

For the experiments involving SMOTE, the pretrained CNN architectures were used exclusively to extract deep feature vectors from the MRI images, while the classifier architecture remained unchanged.

Training was performed using the Adam optimizer with a learning rate of 1×10^{-5} , a batch size of 32 and for 12 epochs.

4. Results

4.1. Evaluation

The performance of the models was measured using the metrics accuracy, precision, recall, and F1-score, defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\text{-score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$Balanced\ Accuracy = \frac{Recall + Specificity}{2} \quad (6)$$

where TP represents true positives, TN true negatives, FP false positives, and FN false negatives.

Additionally, ROC-AUC was computed to measure the overall ability of the models to distinguish between Parkinson's disease and healthy subjects independently of a fixed decision threshold.

In the medical field, reducing the percentage of false negative cases is essential. The lack of a correct diagnosis in the case of the existence of the disease will delay the necessary treatment, which will lead to aggravation of symptoms and permanent damage.

DenseNet121, ResNet50, ResNet101, and VGG16 were trained and tested on three variations of the initial dataset: the unchanged set, the set obtained with CLAHE and the set balanced with SMOTE (Synthetic Minority Oversampling Technique).

Table 3 shows a comparison of the accuracy obtained for each model depending on the dataset and the techniques used.

It is evident that ResNet50 achieved the highest individual accuracy of 95.31% under the CLAHE strategy. After applying SMOTE to the extracted deep feature vectors, ResNet101 obtained the highest accuracy (94.81%), followed by VGG16 (91.11%), whereas DenseNet121 and ResNet50 both reached 89.63%. The ROC-AUC results obtained by the different architectures are presented in Table 4.

On the unbalanced dataset, the models have very poor accuracy, around 50%. Initially, DenseNet121 achieved better performance on the majority class in this experiment, which may be related to its dense connectivity structure, where the current layer receives as input the output of all previous layers. The performance of ResNet50 and ResNet101, which use skip connections and favor fast learning of dominant features, and VGG16, a simpler architecture, increases more drastically due to preprocessing techniques, while DenseNet121 shows signs of overfitting. The low balanced accuracy values depicted in Table 5 show that the models failed to generalize across both classes. A notable difference is observed in the ability to correctly identify healthy subjects. According to Table 6, VGG16 achieved a specificity of 99.56% and ResNet101 achieved 99.12%. The bias toward the

majority class is accentuated by the recall Tables 7–10, where, for example, DenseNet121 achieved a PD recall of 96% while correctly classifying only 20% of healthy subjects.

Table 3. Accuracy obtained by different architectures under different preprocessing strategies.

Method	DenseNet121 (%)	ResNet50 (%)	VGG16 (%)	ResNet101 (%)
Unbalanced	53.33	51.23	57.04	59.14
Unbalanced + CLAHE	93.21	95.31	91.98	94.44
Balanced + CLAHE + SMOTE	89.63	89.63	91.11	94.81

Table 4. ROC-AUC obtained by different architectures under different preprocessing strategies.

Method	DenseNet121 (%)	ResNet50 (%)	VGG16 (%)	ResNet101 (%)
Unbalanced	58.89	56.08	59.98	60.29
Unbalanced + CLAHE	98.87	98.98	98.96	99.34
Balanced + CLAHE + SMOTE	97.04	95.43	97.7	98.76

Table 5. Balanced Accuracy obtained by different architectures under different preprocessing strategies.

Method	DenseNet121 (%)	ResNet50 (%)	VGG16 (%)	ResNet101 (%)
Unbalanced	58.06	53.37	59.84	60.19
Unbalanced + CLAHE	93.86	95.30	90.91	93.79
Balanced + CLAHE + SMOTE	88.57	89.47	90.26	94.7

The impact of the CLAHE technique is evident, which increases the performance by approximately 40%. The contrast of the images was enhanced, making it easier for the models to identify features, leading to an improved representation of the anatomical structures.

The improvement was consistently reflected in the balanced accuracy Table 5, with ResNet50 reaching 95.30%. As shown in Tables 7–10 the recall for the minority class increased significantly. These gains indicate that CLAHE successfully improved the visibility of anatomical features relevant for classification, allowing the convolutional layers to extract the necessary information.

Table 6. Specificity obtained by different architectures under different preprocessing strategies.

Method	DenseNet121 (%)	ResNet50 (%)	VGG16 (%)	ResNet101 (%)
Unbalanced	19.78	36.04	37.14	51.65
Unbalanced + CLAHE	88.57	95.38	99.56	99.12
Balanced + CLAHE + SMOTE	97.14	90.77	97.14	95.6

Table 7. Recall obtained by DenseNet121 under different preprocessing strategies.

Method	Non-PD Recall (%)	PD Recall (%)
Unbalanced	20	96
Unbalanced + CLAHE	89	99
Balanced + CLAHE + SMOTE	97	80

Table 8. Recall obtained by ResNet50 under different preprocessing strategies.

Method	Non-PD Recall (%)	PD Recall (%)
Unbalanced	36	71
Unbalanced + CLAHE	95	95
Balanced + CLAHE + SMOTE	91	88

Table 9. Recall obtained by ResNet101 under different preprocessing strategies.

Method	Non-PD Recall (%)	PD Recall (%)
Unbalanced	52	69
Unbalanced + CLAHE	99	88
Balanced + CLAHE + SMOTE	97	83

Table 10. Recall obtained by VGG16 under different preprocessing strategies.

Method	Non-PD Recall (%)	PD Recall (%)
Unbalanced	37	83
Unbalanced + CLAHE	100	82
Balanced + CLAHE + SMOTE	96	94

As illustrated in Figure 2, a divergence can be observed between the training and validation curves for the VGG16 model when using the unbalanced dataset. The results indicate that initially, the network failed to generalize effectively. However, after applying the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique, both training loss and validation loss decreased. The precision and F1-score results obtained by the evaluated architectures are presented in Tables 11–14 and Tables 15–18, respectively.

Table 11. Precision obtained by DenseNet121 under different preprocessing strategies.

Method	Non-PD Precision (%)	PD Precision (%)
Unbalanced	87	48
Unbalanced + CLAHE	99	87
Balanced + CLAHE + SMOTE	86	96

Table 12. Precision obtained by ResNet50 under different preprocessing strategies.

Method	Non-PD Precision (%)	PD Precision (%)
Unbalanced	61	46
Unbalanced + CLAHE	96	94
Balanced + CLAHE + SMOTE	91	88

Table 13. Precision obtained by ResNet101 under different preprocessing strategies.

Method	Non-PD Precision (%)	PD Precision (%)
Unbalanced	68	53
Unbalanced + CLAHE	92	99
Balanced + CLAHE + SMOTE	95	94

Table 14. Precision obtained by VGG16 under different preprocessing strategies.

Method	Non-PD Precision (%)	PD Precision (%)
Unbalanced	73	51
Unbalanced + CLAHE	88	99
Balanced + CLAHE + SMOTE	88	96

Table 15. F1-score obtained by DenseNet121 under different preprocessing strategies.

Method	Non-PD F1-Score (%)	PD F1-Score (%)
Unbalanced	32	64
Unbalanced + CLAHE	94	93
Balanced + CLAHE + SMOTE	91	87

Table 16. F1-score obtained by ResNet50 under different preprocessing strategies.

Method	Non-PD F1-Score (%)	PD F1-Score (%)
Unbalanced	45	56
Unbalanced + CLAHE	96	95
Balanced + CLAHE + SMOTE	91	88

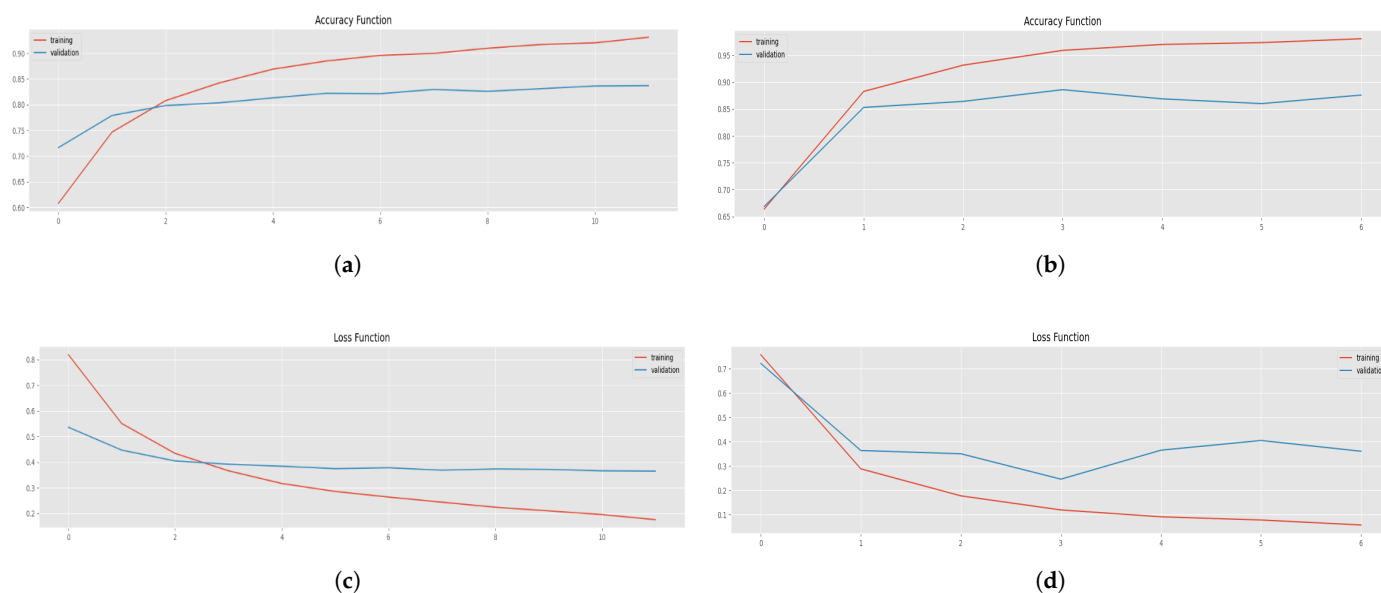
Table 17. F1-score obtained by ResNet101 under different preprocessing strategies.

Method	Non-PD F1-Score (%)	PD F1-Sscore (%)
Unbalanced	59	60
Unbalanced + CLAHE	95	93
Balanced + CLAHE + SMOTE	95	94

Table 18. F1-score obtained by VGG16 under different preprocessing strategies.

Method	Non-PD F1-Score (%)	PD F1-Score (%)
Unbalanced	49	63
Unbalanced + CLAHE	93	90
Balanced + CLAHE + SMOTE	92	89

Adding SMOTE to correct class imbalance yielded mixed results. As shown in Table 3, ResNet101 achieved the highest accuracy (94.81%) on the balanced dataset, followed by VGG16 (91.11%), while DenseNet121 and ResNet50 both obtained 89.63%. Compared with the CLAHE configuration, SMOTE slightly improved the performance of ResNet101, whereas the remaining architectures showed lower accuracies.

**Figure 2.** Training curves obtained with the VGG16 model. (a) Accuracy on the CLAHE with SMOTE dataset. (b) Accuracy on the CLAHE dataset. (c) Loss on the CLAHE with SMOTE dataset. (d) Loss on the CLAHE dataset.

Consistent with the desired results in the medical domain, the most relevant observation concerns the recall of the Parkinson's disease class. VGG16 achieved the highest PD recall of 94% (Table 10), indicating a strong ability to identify Parkinson's disease patients. ResNet50 obtained a PD recall of 88% (Table 8), followed by ResNet101 with 83% (Table 9) and DenseNet121 with 80% (Table 7). These results indicate that the effect of SMOTE

depended on the CNN architecture and did not consistently improve the sensitivity of all models.

The confusion matrices in Figure 3 illustrate the different error distributions produced by the evaluated architectures. Regarding ROC-AUC (Table 4), all models achieved values above 95%, indicating a strong ability to distinguish between Parkinson's disease and healthy subjects. Among the evaluated models, ResNet101 achieved the highest ROC-AUC (98.76%) for the balanced dataset, whereas the highest overall ROC-AUC (99.34%) was obtained using the CLAHE configuration.

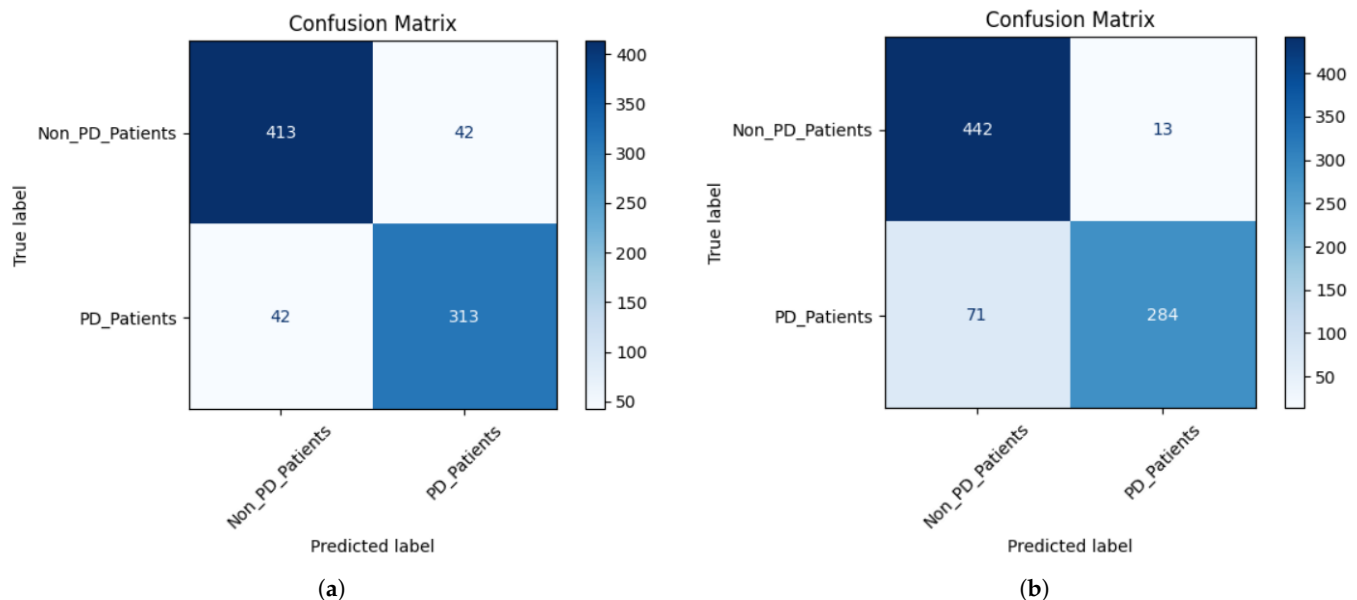


Figure 3. Confusion matrices obtained on the NTUA dataset: (a) ResNet50 and (b) DenseNet121 after applying CLAHE and SMOTE preprocessing.

4.2. Subject-Level Metrics

The patient-level evaluation was performed by aggregating slice predictions per patient. We used mean probability computed from the predicted Parkinson's disease probability of each image. Patients with a mean probability greater than or equal to 0.5 were classified as PD, while those with a mean probability below 0.5 were classified as Non-PD.

For the CLAHE configuration, all six test subjects (four PD and two Non-PD) were correctly classified by all models, corresponding to a patient-level accuracy of 100% (6/6 correctly classified subjects). In the SMOTE-based experiments, ResNet50, VGG16, and ResNet101 also correctly classified all six test subjects. In contrast, DenseNet121 correctly classified four out of six subjects, resulting in a patient-level accuracy of 66.67%. Although the best-performing models achieved perfect patient-level precision, recall, and F1-score for both classes, these results should be interpreted with caution because they were obtained from a limited test cohort of only six subjects.

The class-specific metrics presented in Tables 19 and 20 show differences in the ability of the models to correctly identify non-PD and PD patients in the case of the unbalanced dataset. All models show bias towards the majority class. DenseNet121 demonstrated poor performance for the non-PD class, with a precision, recall, and F1-score of 0%. ResNet50 and ResNet101 achieved identical class-specific results. Among the evaluated architectures, VGG16 achieved the highest patient-level accuracy, reaching 83.33%. It also demonstrated the strongest overall class-specific performance among the evaluated models, with a precision of 100% for Non-PD patients and 80% for PD patients.

Table 19. Accuracy across different CNN architectures utilizing the unbalanced dataset.

Method	DenseNet121 (%)	ResNet50 (%)	VGG16 (%)	ResNet101 (%)
Unbalanced	66.67	66.67	83.33	66.67

Table 20. Patient-level evaluation obtained by the models trained using the unbalanced dataset.

Method	DenseNet121 (%)	ResNet50 (%)	VGG16 (%)	ResNet101 (%)
Precision (Non-PD/PD)	0/67	50/75	100/80	50/75
Recall (Non-PD/PD)	0/100	50/75	50/100	50/75
F1-score (Non-PD/PD)	0/80	50/75	67/89	50/75

4.3. Variability Across Subject-Wise Data Splits

In medical imaging datasets with a limited number of participants, the choice of which subjects are used for testing can substantially influence the observed performance.

An experiment was carried out using the 50-subject axial subset (36 patients, 14 healthy subjects) obtained according to the selection procedure described in Section 3.1. The goal was to reduce irrelevant content from the images and allow the model to focus on brain structures that may contain biomarkers associated with Parkinson’s disease.

Numerous studies in the literature evaluate model performance using a single way of splitting the dataset for training, testing, and validation. The images collected from each patient tend to differ in number and quality. Previous research [16] has demonstrated that assigning a larger and more diverse subset for testing will affect training and lead to underdevelopment of the models due to the lack of useful information.

The proposed approach was implemented using 6 runs with a different distribution. For each split, all MRI slices belonging to a subject were assigned exclusively to the training and validation set or to the testing set.

The results presented in Table 21 demonstrate significant differences between the various subject-wise splits. The impact was especially evident for the healthy class, which was underrepresented compared to the Parkinson’s disease class. The highest accuracy for the DenseNet121 model (85.8%) was obtained in the split containing Subjects 54 and 56 in the test set, while the lowest accuracy (49.19%) was observed in the split containing Subjects 1 and 16. These subjects contributed 92 and 114 images, and their absence from the training set strongly affected the model’s performance. These observations suggest that the limited number of healthy subjects can significantly affect the model’s ability to generalize.

Table 21. Impact of healthy subjects included in the testing set on DenseNet121 performance.

Healthy Subject 1	Healthy Subject 2	No. Images	Accuracy (%)	Precision (Non-PD/PD) (%)	Recall (Non-PD/PD) (%)
S60	S1	19/92	77.00	20/90	32/83
S58	S15	8/17	76.00	15/98	68/77
S16	S1	114/92	49.19	21/55	8/79
S19	S58	17/8	61.22	18/84	36/66
S54	S56	23/23	85.80	41/99	93/85
S2	S54	77/23	65.50	29/88	60/67

Tables 22 and 23 show that VGG16 provides the most consistent performance in terms of Parkinson’s disease recall. It reached an average value of approximately 93.8%, compared to DenseNet121 (76%) and ResNet50 (55.5%). Also, the results suggest a tendency to favor the PD class, as the performance metrics for healthy subjects were generally lower.

Table 24 shows the standard deviation for the models in the experiments performed. The most stable model for Parkinson’s detection is VGG16, as it has the highest average accuracy (81.32%) and the lowest standard deviation for PD recall (5.42%). On the other hand, instability is present in ResNet50 through a high standard deviation for accuracy (28.33%) and values of 0 in some splits.

Table 22. Impact of healthy subjects included in the testing set on ResNet50 performance.

Healthy Subject 1	Healthy Subject 2	No. Images	Accuracy (%)	Precision (Non-PD/PD) (%)	Recall (Non-PD/PD) (%)
S60	S1	19/92	53.61	16/93	69/52
S58	S15	8/17	94.36	0/94	0/100
S16	S1	114/92	65.32	56/78	78/57
S19	S58	17/8	70.75	22/84	28/30
S54	S56	23/23	93.75	63/99	90/94
S2	S54	77/23	18.45	18/0	100/0

Table 23. Impact of healthy subjects included in the testing set on VGG16 performance.

Healthy Subject 1	Healthy Subject 2	No. Images	Accuracy (%)	Precision (Non-PD/PD) (%)	Recall (Non-PD/PD) (%)
S60	S1	19/92	79.09	10/88	10/88
S58	S15	8/17	89.39	19/96	28/93
S16	S1	114/92	60.28	67/60	9/97
S19	S58	17/8	79.59	41/88	44/87
S54	S56	23/23	98.01	85/100	97/98
S2	S54	77/23	81.55	0/82	0/100

Table 24. Aggregate performance comparison across all models over the six subject-wise splits (Mean $\pm$ Standard Deviation).

Model	Accuracy (%)	Precision (Non-PD/PD) (%)	Recall (Non-PD/PD) (%)
DenseNet121	69.12 $\pm$ 13.12	24.00 $\pm$ 9.55/85.67 $\pm$ 16.11	49.50 $\pm$ 30.18/76.17 $\pm$ 8.01
ResNet50	66.04 $\pm$ 28.33	29.17 $\pm$ 24.76/74.67 $\pm$ 37.35	60.83 $\pm$ 38.79/55.50 $\pm$ 37.97
VGG16	81.32 $\pm$ 12.60	37.00 $\pm$ 33.59/85.67 $\pm$ 14.11	31.33 $\pm$ 35.83/93.83 $\pm$ 5.42

However, the experimental results indicate that no single architecture consistently outperformed the others across all evaluation scenarios. DenseNet121 demonstrated greater robustness, as it never reached a recall of 0% for either class. In contrast, ResNet50 achieved high performance only on specific splits and appeared to be the most sensitive to the particular subjects included in the training and test sets. VGG16, on the other hand, produced the highest performance in the best-performing experiment, reaching 98.01% accuracy and 97% recall for the healthy class.

Figure 4 compares the learning curves obtained for the best and worst performing splits for the DenseNet121 model. The worst split showed an increasing validation loss, despite the decreasing training losses, indicating overfitting. On the other hand, the best-performing split exhibited enhanced generalization capability, as both training and validation losses follow a similar decreasing trend.

Since all experiments used the same architecture and hyperparameters, the observed differences are mainly related to the splitting of the dataset at the patient level. The results suggest that the contribution of a subject is not only determined by the number of images available. For example, runs with different splits achieved similar recall values for the healthy class, even though the tables show differences in the number of healthy training

images. This indicates that certain subjects may be more representative of their class, while others bring additional variability that influences the learning process.

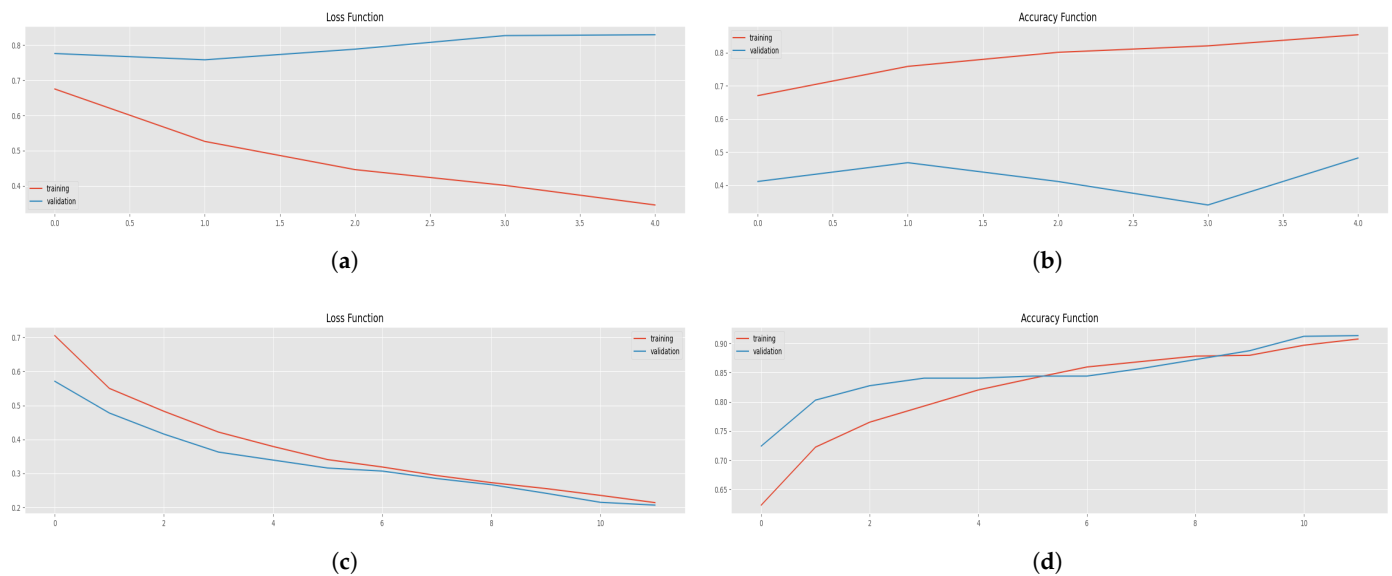


Figure 4. Comparison of the learning dynamics of DenseNet121, obtained for the worst-performing and best-performing subject-wise splits. (a) Training and validation loss for the worst-performing split. (b) Training and validation accuracy for the worst-performing split. (c) Training and validation loss for the best-performing split. (d) Training and validation accuracy for the best-performing split.

4.4. Grad-CAM Analysis

An explainable AI approach based on Grad-CAM was applied to validate the classification results and increase confidence in the predictions provided by the models. It was chosen over other methods such as LIME and SHAP because it is faster, widely used in medical imaging studies, and can be directly applied to convolutional feature maps generated by CNN models.

The axial, coronal, and sagittal Grad-CAM visualizations presented in this section were generated from the complete multi-plane MRI dataset and are included only to illustrate the explainability behaviour of the proposed models. They are not part of the subject-level evaluation reported in Section 4.3, for which the quantitative results correspond exclusively to the axial T1 MRI subset.

Visual explanations help reveal how deep learning models reach their predictions, making it easier to identify potential biases caused by irrelevant image patterns and increasing the transparency of the classification process. In this study, XAI was used to verify whether the models focused on relevant brain structures in MRI scans rather than on imaging artifacts or non-informative anatomical regions such as the face, the neck and the upper part of the skull.

Gradient-weighted Class Activation Mapping (Grad-CAM) was created to preserve both interpretability and faithfulness to the original model. It generates class-discriminative activation maps and highlights the regions in the image that most influenced the model's prediction. The method uses gradient information propagated back to the last convolutional layer to estimate the importance of each region in the image. Warm colors emphasize the dominant regions in the model's decision, while cool colors emphasize the least relevant ones [17].

Models should focus on consistent medical characteristics that signal the presence of Parkinson's disease. Previous research has demonstrated that the important regions are the basal ganglia (especially the putamen and pallidum), the amygdala, the hippocampus,

the brainstem, and the fronto-temporo-parietal cortical areas [18]. Figure 5 shows the application of Grad-CAM on several MRI slices correctly classified by the ResNet50 model. As can be observed from the warm colors, increased attention is paid to the midbrain region. The behavior of the model is consistent with studies that identify brainstem and mesencephalic structures as important regions in Parkinson's disease.

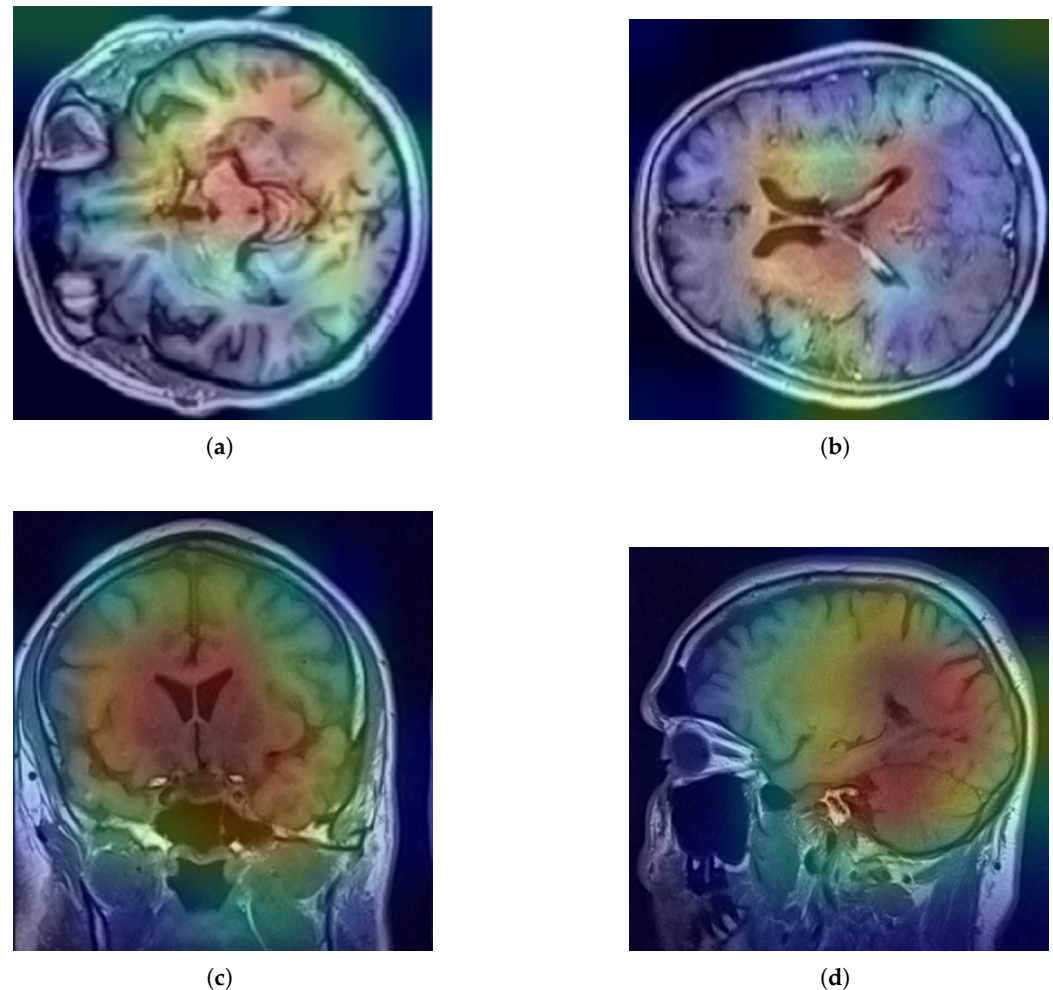


Figure 5. Grad-CAM visualizations generated by the ResNet50 model for correctly classified MRI slices. (a) Subject35 (axial slice), classified as Parkinson's disease with a probability of 0.63. (b) Subject54 (axial slice), classified as healthy with a probability of 0.92. (c) Subject40 (coronal slice), classified as Parkinson's disease with a probability of 1.00. (d) Subject40 (sagittal slice), classified as Parkinson's disease with a probability of 1.00.

Figure 6 highlights the differences between the activation maps obtained by the VGG16 and DenseNet121 models. Both models tend to focus on a single dominant region, although some of their attention is also directed towards areas that may contain limited diagnostic information, such as the eyes or the edges of the brain. In contrast, ResNet50 simultaneously considers multiple central brain structures. Its predictions are based on a broader anatomical context, while the other models generate more localized activation patterns. The findings are consistent with previous studies suggesting that Parkinson's disease is associated with structural changes in multiple brain regions. Models that achieve better performance may benefit from considering information from several relevant areas, including deep gray matter structures and frontotemporal regions, rather than focusing on a single point of interest [19]. A possible explanation for the observed differences may be the architecture of each neural network. VGG16 tends to compress image information through three fully connected layers that eliminate distributed spatial information as it approaches

the classification step [15], whereas ResNet50 maintains richer feature representations through skip connections and global average pooling [14]. It must be noted that some activation patterns extend toward non-diagnostic peripheral structures, such as the eyes and image borders, indicating localized model biases. Consequently, this Grad-CAM analysis serves as a qualitative exploratory tool for transparency rather than a definitive biological validation.

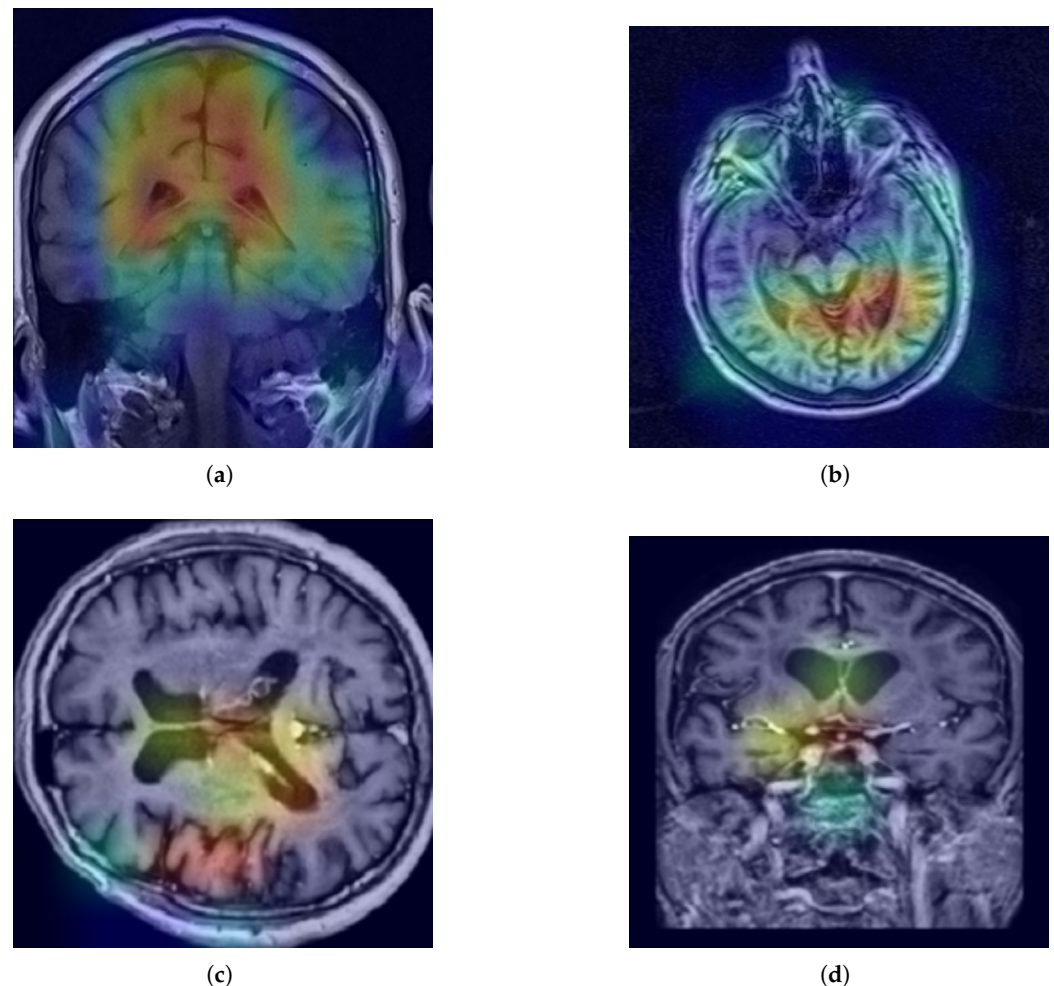


Figure 6. Grad-CAM visualizations generated by the VGG16 and DenseNet121 models for correctly classified MRI slices. (a) VGG16: Subject40 (coronal slice), classified as Parkinson's disease with a probability of 1.00. (b) VGG16: Subject19 (axial slice), classified as healthy with a probability of 1.00.

(c) DenseNet121: Subject25 (axial slice), classified as Parkinson's disease with a probability of 0.88. (d) DenseNet121: Subject6 (coronal slice), classified as Parkinson's disease with a probability of 1.00.

4.5. Comparison with Previous Work

Table 25 shows a comparison of the highest PD recall obtained by the evaluated models with medical studies that used the NTUA Parkinson's Dataset. It is observed that the performance of DenseNet121 (99%) aligns with the results of other studies such as PDCNet (97.69%) and CNN-RNN (96%). ResNet50 also achieved a high PD recall of 95%, while VGG16 reached 94% and ResNet101 achieved 88%. The performance of the developed method compared with existing state-of-the-art approaches is presented in Table 26. Overall, the proposed framework provides competitive Parkinson's disease detection performance compared with previously reported approaches on the same dataset.

Table 25. Comparison of PD recall values reported on the NTUA dataset.

Method	PD Recall (%)
CNN-RNN [7]	96
PDCNet (dropout = 0.3) [9]	97.69
DenseNet121 + CLAHE	99
ResNet101 + CLAHE	88
VGG16 + CLAHE	94
ResNet50 + CLAHE	95

Table 26. Accuracy comparison of state-of-the-art methods evaluated on the NTUA dataset.

Method	Accuracy (%)
CNN [7]	94.00
CNN-RNN [7]	98.00
DenseNet121 [8]	90.70
MobileNet V3 [8]	96.53
Optimized MobileNet V3 [8]	98.36
ResNet50 + CSAM [8]	99.16
PDCNet (dropout = 0.1) [9]	96.23
PDCNet (dropout = 0.2) [9]	97.56
PDCNet (dropout = 0.3) [9]	98.27
PDCNet (dropout = 0.4) [9]	97.85
DenseNet121 (Balanced + CLAHE)	93.21
ResNet50 (Unbalanced + CLAHE)	95.31
VGG16 (Balanced + CLAHE)	91.98
ResNet101 (Balanced + CLAHE + SMOTE)	94.81

Nevertheless, due to variations in experimental protocols and imaging modalities, comparisons with prior works in the literature are purely indicative rather than strictly equivalent.

4.6. Limitations and Future Work

Current limitations include training on a small dataset, only 78 subjects, and the use of a single type of medical imaging, MRI Axial T1 sequence. The latter focuses only on a few features resulting from the impact of Parkinson's disease on the human brain. NTUA Parkinson's Dataset also contains images obtained through DTI imaging, which is useful for white matter integrity and microstructural changes, and T2-weighted MRI, emphasizing lesions. In addition, many of the studies in the literature use a combination of different models and datasets. So future work will focus on validating these findings on larger and more diverse multi-center datasets in order to further improve robustness, generalization capability, and real-world clinical applicability.

5. Discussion

Among the evaluated models, ResNet50 achieved the highest overall accuracy of 95.31%, showing strong performance in the evaluated dataset after preprocessing.

From a clinical perspective, DenseNet121 proved particularly promising due to its superior sensitivity, achieving a recall of 99% for the Parkinson's disease class and therefore reducing the number of false negatives, which is critical in medical diagnosis, though the inclusion of SMOTE demonstrated mixed effects on overall training stability. On the other hand, the models achieved 100% accuracy on the CLAHE and SMOTE datasets when evaluated at the patient level.

Furthermore, to improve model interpretability and clinical transparency, Grad-CAM visualizations were also generated, highlighting the brain regions that contributed most strongly to the final predictions as part of a qualitative exploratory analysis.

6. Conclusions

This paper investigated deep learning approaches for the automated diagnosis of Parkinson's disease using T1-weighted MRI scans, with a comparative analysis of the ResNet50, DenseNet121, and VGG16 architectures. To improve image quality and model robustness, CLAHE was applied for contrast enhancement, while SMOTE was employed to address dataset imbalance within the NTUA Parkinson's dataset. Beyond conventional classification experiments, this study introduced a comprehensive evaluation framework that included multiple patient-wise train-test splits in order to assess model generalization and robustness across subjects. In addition, a targeted analysis using only axial MRI slices was conducted to reduce irrelevant anatomical information and emphasize brain regions potentially associated with Parkinson's disease. To improve model interpretability and clinical transparency, Grad-CAM visualizations were also generated, highlighting the brain regions that contributed most strongly to the final predictions as part of a qualitative exploratory analysis.

The obtained results demonstrate that adaptive preprocessing, explainable artificial intelligence, and patient-wise evaluation provide a useful framework for MRI-based Parkinson's disease detection, although the effectiveness of SMOTE depends on the CNN architecture and does not consistently improve performance across all evaluated models.

As a preliminary, dataset-specific evaluation, the results highlight that performance variability can be highly dependent on subject selection. The proposed framework contributes toward the development of more interpretable and reproducible computer-aided diagnostic research. The goal of this study has not been to recommend a single model, but to underline that each approach can be appropriate depending on the specific characteristics of the dataset used.

Author Contributions: The concept of the article was proposed by N.P., the data resources and validation were contributed by I.-T.I., the formal analysis, investigation, and draft preparation were performed by I.-T.I. The supervision and review of the study were headed by N.P. The final writing was critically revised by N.P. and finally approved by the authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the project “Romanian Hub for Artificial Intelligence—HRIA”, Smart Growth, Digitization and Financial Instruments Program, 2021–2027, MySMIS no. 351416.

Data Availability Statement: Data is contained within the article, further inquiries can be directed to the corresponding author.

Acknowledgments: This research was supported by the project “Romanian Hub for Artificial Intelligence—HRIA”, Smart Growth, Digitization and Financial Instruments Program, 2021–2027, MySMIS no. 351416.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Poewe, W.; Seppi, K.; Tanner, C.M.; Halliday, G.M.; Brundin, P.; Volkmann, J.; Schrag, A.E.; Lang, A.E. Parkinson disease. *Nat. Rev. Dis. Prim.* **2017**, *3*, 17013.
2. Jokinen, P.; Brück, A.; Aalto, S.; Forsback, S.; Parkkola, R.; Rinne, J.O. Impaired cognitive performance in Parkinson's disease is related to caudate dopaminergic hypofunction and hippocampal atrophy. *Park. Relat. Disord.* **2009**, *15*, 88–93. [[CrossRef](#)] [[PubMed](#)]

3. Khanna, K.; Gambhir, S.; Gambhir, M. Current challenges in detection of Parkinson's disease. *J. Crit. Rev.* **2020**, *7*, 1461–1467. [[CrossRef](#)] [[PubMed](#)]
4. Noor, M.B.T.; Zenia, N.Z.; Kaiser, M.S.; Mamun, S.A.; Mahmud, M. Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. *Brain Inform.* **2020**, *7*, 11. [[CrossRef](#)] [[PubMed](#)]
5. Uzzaman, A.; Sakib, A.N.; Shushmita, S.A.; Kabir, S.A.; Reza, M.T.; Parvez, M.Z. Parkinson's disease detection using FMRI images leveraging transfer learning on convolutional neural network. In Proceedings of the 2020 International Conference on Machine Learning and Cybernetics (ICMLC), Adelaide, Australia, 2 December 2020; pp. 131–136.
6. Kaur, S.; Aggarwal, H.; Rani, R. Diagnosis of Parkinson's disease using deep CNN with transfer learning and data augmentation. *Multimed. Tools Appl.* **2021**, *80*, 10113–10139. [[CrossRef](#)]
7. Tagaris, A.; Kollias, D.; Stafylopatis, A.; Tagaris, G.; Kollias, S. Machine learning for neurodegenerative disorder diagnosis—survey of practices and launch of benchmark dataset. *Int. J. Artif. Intell. Tools* **2018**, *27*, 1850011. [[CrossRef](#)]
8. Anuranjani, K.; Karthi, A. Optimizing Parkinson's disease Classification in MRI Scans Using a Modified ResNet50 Architecture. *CLEI Electron. J.* **2025**, *28*, 9. [[CrossRef](#)]
9. Prabakaran, T.; Latha, R.; Kavitha, R.; Joshi, S.; Radhika, M.; Srinivasan, C. PDCNet: Parkinson's Disease Classification Network using MRI Scans. *J. Innov. Image Process.* **2025**, *7*, 207–225. [[CrossRef](#)]
10. Chakraborty, S.; Aich, S.; Kim, H.C. Detection of Parkinson's disease from 3T T1 weighted MRI scans using 3D convolutional neural network. *Diagnostics* **2020**, *10*, 402. [[CrossRef](#)] [[PubMed](#)]
11. Zuiderveld, K. Contrast Limited Adaptive Histogram Equalization. In *Graphics Gems IV*; Heckbert, P.S., Ed.; Academic Press: Boston, MA, USA, 1994; pp. 474–485.
12. Fernández, A.; Garcia, S.; Herrera, F.; Chawla, N.V. SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *J. Artif. Intell. Res.* **2018**, *61*, 863–905. [[CrossRef](#)]
13. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
15. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
16. Steyerberg, E.W.; Harrell, F.E., Jr. Prediction models need appropriate internal, internal-external, and external validation. *J. Clin. Epidemiol.* **2016**, *69*, 245–247. [[CrossRef](#)] [[PubMed](#)]
17. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
18. Veetil, I.K.; Chowdary, D.E.; Chowdary, P.N.; Sowmya, V.; Gopalakrishnan, E. An analysis of data leakage and generalizability in MRI based classification of Parkinson's Disease using explainable 2D Convolutional Neural Networks. *Digit. Signal Process.* **2024**, *147*, 104407. [[CrossRef](#)]
19. Datta, D.; Debnath, T.; Chau, M.; Paul, M.; Adhikary, G.; Rahman, M.G. A Sparse-Attention Deep Learning Model Integrating Heterogeneous Multimodal Features for Parkinson's Disease Severity Profiling. *arXiv* **2026**, arXiv:2601.00519.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.