



Advanced Parkinson's Disease Detection: A comprehensive artificial intelligence approach utilizing clinical assessment and neuroimaging samples

Nusrat Islam, Md. Shaiful Alam Turza, Shazzadul Islam Fahim, Rashedur M. Rahman *

Department of Electrical and Computer Engineering, North South University, Dhaka, 1229, Bangladesh



ARTICLE INFO

Keywords:

Machine learning
Explainable artificial intelligence
Local interpretable model agnostic explanations
Shapley additive explanations

ABSTRACT

Medical experts are utilizing neuroimaging and clinical assessments to enhance the early identification of Parkinson's disease. The current research initiative offers ways to identify Parkinson's disease using machine learning and transfer learning. To carry out this, we extracted 7500 MRI images from 2022 and 2023 and 12 clinical assessment records from 2010 to 2023 from the well-known Parkinson's Progression Marker Initiative (PPMI) database. Then, we applied machine and transfer learning approaches using clinical assessment records and MRI images, respectively. To identify Parkinson's Disease (PD) using samples from clinical assessments, four distinct resampling techniques were employed. Subsequently, three machine learning models were applied to train on these resample records, and the recall score was analyzed. A hybrid of SMOTE and ENN proved to be the most effective approach for handling all of the imbalanced data, according to the recall study. Later, four different feature selection methods were used to find the top 10 features using these new samples. Lastly, we trained and validated the model using nine machine-learning algorithms. We also used explainable AI techniques like LIME and SHAP to interpret clinical assessment records. The extra tree classifier outperformed the others in terms of accuracy, reaching 98.44% using the tree-based feature selection technique. In addition to examining clinical assessment samples, this study investigated Parkinson's disease using neuroimaging data. In pursuit of this objective, four pre-trained architectures were employed to analyze MRI images through two distinct approaches. The first approach involved utilizing the convolutional layer while replacing the remaining two layers with a customized Artificial Neural Network (ANN). Subsequently, training and evaluation are performed using our MRI samples, followed by analyzing significant weights using a LIME interpretable explainer. The second approach employs an improvisational technique without directly replacing the last layer. Instead, we predicted the weights of our MRI samples using the knowledge of the pre-trained model and stored them. Finally, CNN architecture was utilized for Parkinson's disease detection, achieving an optimal accuracy of 85.08% with the implementation of DenseNet169 and CNN.

1. Introduction

Parkinson's disease is an advancing neurological condition ailment defined by the slow decline of brain neurons that produce dopamine, a significant neurotransmitter important for controlling movement (Bereczki, 2010). This disability interferes with the normal functioning of the nervous system, affecting the organs these nerves regulate. The consequences extend beyond movement-related challenges, encompassing various non-motor symptoms such as disruptions in voice and speech, disability to smell or detect odors, depression, Autonomic Dysfunction, sleep disturbances, and cognitive impairment (Ahmad, Longhurst, Stiles, Downard, & Martin, 2023). Additionally, individuals with Parkinson's may experience difficulties

in synchronizing eye movements and controlling their motion (Gupta et al., 2021), further contributing to the complex array of symptoms associated with this debilitating disease. Parkinson's disease was first formally described by an English physician named James Parkinson. In 1987, He published a detailed essay that provided a comprehensive description of the symptoms which he referred to as "shaking palsy" to highlight the characteristic tremors associated with the disorder. On top of that, He identified not only the characteristic tremors but also the slowness of movement, muscle rigidity, and postural instability, which are now recognized as key features of Parkinson's disease (Parkinson, 2002).

* Corresponding author.

E-mail address: rashedur.rahman@northsouth.edu (R.M. Rahman).

Parkinson's disease has received extensive and ongoing attention from the medical community, which has led to a growing interest in understanding and treating this difficult illness. Although some cases start before the age of 50, Parkinson's disease usually strikes people around the age of 70. Parkinson's disease tends to strike more persons after the age of sixty. It is interesting to note that the number of Parkinson's patients rises sharply as people age, particularly beyond 60 (Gazewood, Richards, & Clebak, 2013). This indicates that it gets more prevalent in older age groups. Additionally, these studies offer valuable evidence regarding the efficacy of different therapeutic interventions. Conversely, neuroimaging techniques, encompassing MRI, SPECT, and PET (Dabrowska, Schinwelski, Sitek, Muraszko-Klaudel, Brockhuis, Jamrozik, & Slawek, 2015), offer invaluable insights into the underlying pathology within the brain. These technologies enable medical professionals to uncover structural and functional changes, aiding in diagnosing neurodegenerative diseases, including PD. Additionally, neuroimaging is pivotal in advancing research and understanding of these complex neurological conditions.

Many prior research studies have predominantly concentrated on singular analytical methods for detecting Parkinson's disease. This singular focus often involves utilizing clinical assessments, such as surveys, MRI, SPECT, and PET scans of the brain, or the examination of voice samples. However, this approach poses a limitation as it lacks a comprehensive and detailed investigation across these diverse areas. Consequently, there is a risk of missing a nuanced understanding of Parkinson's disease by not integrating information from various sources. To enhance our comprehension, it is crucial to integrate insights derived from multiple sources. This multidimensional approach can facilitate a more thorough and holistic analysis of Parkinson's disease, potentially uncovering deeper insights into its complexities. Therefore, to overcome the constraints, we are performing an analysis that combines clinical evaluations with neuroimaging to provide a holistic picture. As a result, we can comprehend a wide range of subjective viewpoints as well as intricate details about the composition and operation of the brain. Clinical assessments aid (Evangelou, Tsianos, & Ioannidis, 2008) in comprehending a patient's symptoms, medical history, and lifestyle factors. Moreover, these assessments explore psychosocial aspects, evaluating the patient's mental and emotional well-being, along with their overall quality of life. Uncovering patient preferences, treatment expectations, and adherence to medical advice, these evaluations provide a holistic understanding. Conversely, MRI plays a crucial role in identifying structural abnormalities such as tumors, vascular irregularities, and developmental anomalies in the brain Yen, Lin, and Chiang (2023). Furthermore, it proves invaluable for assessing brain function, especially with advanced techniques like functional MRI, which measures changes in blood flow and assists in mapping brain activity. Thus, this research aims to develop a machine and transfer learning technique by analyzing patient data from clinical assessments and neuroimaging that can determine if a patient has Parkinson's disease and assess the key factors that contribute to the issue. The following is a summary of the study's main contributions:

- 12 distinct sets of clinical assessment records from the years between 2010 to 2023 and a total of 7500 MRI images from the years 2022 and 2023 are collected from a reputable database named Parkinson's Progression Marker Initiative (PPMI). This global research project was started in 2010, and it involves co-operation between academics, practitioners of medicine, and patients. Its main objective is to create an extensive dataset and a collection of biological samples to advance Parkinson's disease research.
- This study applied to machine and transfer learning approaches using both clinical assessment records and MRI images, respectively. We use four distinct resampling approaches to carry out the machine learning process, and from there, we determine that the hybrid of SMOTE and ENN is the most effective way to

manage imbalanced data. Following that, these new samples are used to determine the top 10 features using four distinct feature selection techniques. The model is trained and validated using nine machine-learning techniques. Utilizing the tree-based feature selection technique, the Extra tree classifier exceeds the others in terms of accuracy, attaining 98.44%.

- Explainable AI techniques i.e., LIME and SHAP are used to interpret our model clinical assessment records.
- This study also adopts four transfer learning pre-trained architectures to analyze the MRI images. These models are already pre-trained with a large dataset named Imagenet. To carry out this, we applied two approaches. For the first approach, we used the convolutional layer while replacing the remaining two layers with our own customized Artificial Neural Network (ANN) and then trained and evaluated our MRI samples. Afterward, we analyzed the important weights of MRI images using a LIME interpretable explainer. For the second approach which involved improvisation, without replacing the last layer directly, we predicted the weights of our MRI samples using the knowledge of the pre-trained model and stored them. Lastly, we utilized CNN architecture to fit these features to detect PD. We have achieved the best accuracy of 85.08% using DenseNet169.

To the best of our knowledge, we are the first to acquire a novel dataset consisting of 7500 MRI images collected between 2022 and 2023, in addition to 12 clinical assessment records spanning from 2010 to 2023, sourced from the Parkinson's Progression Marker Initiative (PPMI) database. Our study is distinctive in its exploration of various data types, integrating clinical assessments and MRI samples, and employing a combination of machine learning and transfer learning techniques. Notably, we prioritize Explainable AI by utilizing LIME and SHAP for interpreting clinical assessments. Additionally, LIME also aids in comprehending the significant predictive factors of MRI samples, thereby enhancing model transparency.

The remaining portion of the document is organized as follows. Section 2 covers the background research on our topic and describes how the field's expertise has evolved. Section 3 presents our overall system design and methodology, the data processing pipeline selection, PD identification utilizing clinical assessments and neuroimaging approaches, and XAI application. After that, Section 4 provides a thorough explanation of the experiment's methods, testing outcomes, and essential characteristics of the clinical assessments and neuroimaging datasets. Section 6 provides an overview of the study's future scope and serves as a conclusion to Section 5, which covers the study's key components.

2. Related works

Advancements in machine learning and data analytics have accelerated the development of non-intrusive patient monitoring techniques, enhancing healthcare monitoring and diagnostics efficiently.

Pavan Rajkumar and his colleagues (Magesh, Myloth, & Tom, 2020) proposed a logical justification for the finding as well as a machine learning algorithm that can reliably predict if a DaTscan SPECT image indicates Parkinson's disease or not. The data used in the model comes from the PPMI database, which comprises 642 SPECT scans divided into PD and non-PD groups. The model's accuracy was 95.2% using transfer learning and the VGG16 CNN architecture. The project aimed to improve early Parkinson's disease identification by making it quicker and easier to utilize in practical situations. Larger datasets and less acute class discrepancies may be used in future studies. Input limitations on neural networks may make it possible to train a whole 3D volume image, and model accuracy improvements may reduce the number of false positives and negatives.

R. Prashanth and his fellow mates (Prashanth, Roy, Mandal, & Ghosh, 2016) created a technique that used Naive Bayes Classifiers

(NB), Support Vector Machine (SVM), Random Forests classifiers, and Boosted Trees to distinguish early Parkinson's disease (PD) subjects from normal using non-motor features like RBD and olfactory loss, as well as biomarkers like cerebrospinal fluid measurements and dopaminergic imaging markers. The SVM classifier performed exceptionally well, with an astounding accuracy rate of 96.40%. With a specificity of 95.01%, it is highly proficient in accurately detecting positive cases and correctly identifying negative ones with a sensitivity of 97.03%. Moreover, an impressive 98.88% area under the ROC curve highlights the model's overall discriminative capabilities. To ensure the model's robustness and generalization, the dataset was carefully partitioned, with 70% of the data set aside for training and the remaining 30% for rigorous testing.

Khushal Thakur and his colleagues ([Thakur, Kapoor, Singh, Sharma, & Malhotra, 2022](#)) developed an effective machine learning model based on voice features for diagnosing Parkinson's disease. SVM, Decision Trees Classifier, Extra Trees Classifier, and Random Forest Classifier were the four assessed models. When used in conjunction with the SelectFromModel feature selection strategy, the Extra Trees Classifier achieved the greatest accuracy of 93.39%. The accuracy of the Random Forest Classifier, which was 91.57%, was also good. The dataset comprised 756 instances, each characterized by 754 attributes. It encompassed data from 64 individuals without health concerns and 188 individuals diagnosed with Parkinson's disease. After using the SelectFromModel feature selection strategies, the models' performance metrics increased. The Extra Trees Classifier proved to be the most effective model for predicting Parkinson's disease using the confusion matrix, with seven (7) healthy participants and 99 patients indicating positive cases.

Arti Rana and her fellow mates ([Rana et al., 2022](#)) developed a model that identified Parkinson's disease (PD) from speech recordings using the suggested machine learning model, which employs SVM, Naive Bayes, ANN, and KNN algorithms. The best classifier is ANN, and the model is computationally efficient. 195 individuals, 147 of whom had Parkinson's disease, provided the dataset for the study, which comprised a collection of acoustic speech measurements. In comparison to SVM and naive Bayes classifiers, the model had an accuracy of 87.17%. To enhance the performance of classifiers, the model can be applied to various machine-learning techniques and datasets.

Cağatay Berke Erdas and his colleagues ([Erdaş & Sümer, 2023](#)) applied 2D & 3D MRI analyses to perform 2D and 3D convolutional neural networks, correlation coefficient (R), coefficient of determination (R² score), Mean Absolute Error (MAE), Median Absolute Error (MedAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) to detect and predict Parkinson's disease (PD). They used 1130 image samples in total, 259 of which were from people without health problems and 871 of which were from people with Parkinson's disease. The image samples were taken from the PPMI dataset. An image registration technique was used to improve the analysis, allowing orientation parameters and coordinates on a fixed image to be determined. This study was isolated into two groups Parkinson's disease identification, a classification problem, and Parkinson's disease severity prediction, which can be viewed as a regression problem. They got the best result in Median Slices, which is 96.20%.

Mohanad Arafe and his colleagues, as documented in their referenced paper ([Arafe et al., 2023](#)), aimed to replicate and reconstruct the procedure for constructing a machine learning model predicting Parkinson's disease using T1-weighted MRI samples. They utilized the Parkinson's Progress Markers Initiative dataset, adhering to the original methods, and examining slight variations in cohort selection, ML model design, feature extraction, and evaluation techniques. They used the seven best characteristics from segmented WM masks of Parkinson's disease patients to train a linear Support Vector Machine (SVM) and then compared it with other machine learning techniques. They created an SRC cohort, consisting of 72 progressive and 72 stable patients scanned using Siemens MRI machines, which has a similar age and F/M

balance as the previous study but differs by baseline HYS values. Their model closely reproduced the original study, achieving a peak AUC of 0.685, better than chance performance but lower than the AUC value of 0.795.

Milton Camacho and his colleagues ([Camacho et al., 2023](#)) worked to develop and evaluate a robust, explainable deep-learning model for Parkinson's disease classification. They applied CNNJacobians, CN-NCombined, CNNIntensity, and SVM-RBF Parkinson to perform the work. They collected data from 13 different resources, which are PPMI, COMPASS-ND, BioCog, PD-MCI Calgary, C-Big, NEUROCON, Tao Wu, ON Japan, Hamburg, UK Biobank, OASIS3, SALD, PD-MCI Montreal. A total of 2041 samples were collected to perform the work. To improve the performance of the fundamental fully convolutional neural (SFCN) network, this study used deep learning approaches. Before this, the SFCN network has proven to be a pioneer in classifying sex with cutting-edge precision and predicting adult mental age, especially when applied to T1-weighted MRI datasets. They got the best result in the CNN Jacobians model, which is 79.3%.

Yida Wang and his fellow mates ([Wang et al., 2023](#)) proposed a convolutional neural network-based DL pipeline for autonomous Parkinson's disease diagnosis and SE-ResNeXt50 models to segment brain nuclei from QSM and T1 W images, accurately distinguishing PD from Healthy Controls (HC). The research utilized two separate datasets: Dataset 1, which included 287 Healthy Controls (HC) and 92 Parkinson's disease (PD) patients from Ruijin Hospital, and Dataset 2, which included 83 PD patients and 72 HC individuals from Zhengzhou University's First Affiliated Hospital. When it came to brain nuclei segmentation, the CA-Net model performed admirably, producing reliable findings and outperforming hand drawings in terms of accuracy. However, some nuclei may show visual differences. The AG-SE-ResNeXt50 model achieved 92.0% accuracy, while the diagnostic model tested with manually segmented ROIs achieved 98.7%.

Iswarya Kannoth Veetil and her colleague ([Veetil, Gopalakrishnan, Sowmya, & Soman, 2021](#)) evaluate five deep learning architectures for improving Magnetic Resonance Imaging (MRI)-based diagnosis of Parkinson's disease, finding three architectures that significantly improve performance, aiding in the selection of a suitable DNN model. They utilized the Parkinson Progression Marker Initiative database to analyze Axial T2 MRI images. The initial data showed poor training accuracy due to class imbalance. To balance, 20,000 images were added, with 10,000 in each class. The experiment used an 80–20 training–testing split, minimal pre-processing, and a Gaussian filter. Three measures of network performance are consistent, with VGG models showing the highest accuracy. VGG19 outperforms VGG16 with 92.60% accuracy and F1 scores for PD and NC classes, while AlexNet achieves 88.90% accuracy.

Puppala PRANEETH and his colleague ([Praneeth et al., 2023](#)) introduced a novel method for categorizing Parkinson's disease using MRI brain scans. They have collected image data from the Parkinson's Progression Marker Initiative (PPMI). Researchers introduced a dataset where 179 patients served as control patients, while 412 patients had been diagnosed with Parkinson's disease (PD). The data was pre-processed using a median filter to remove noise and the min–max normalization method for normalization. The Binary Dragonfly method was employed for feature selection. After that, the study examined the effectiveness of other methods, such as DMVDA (20), DNN (21), SVM (18), and DRCNN, in order to ascertain which strategy produced the best outcomes. They got two results without optimization and with optimization. In both results, they got the best result in Using DRCNN. Without optimization, They got 97.22% accuracy, and with optimization, They got 98.87% accuracy.

Sanjana Tomer and her fellow mate's ([Tomer, Khanna, Gambhir, & Gambhir, 2022](#)) proposed a technique where MRI brain pictures from the PPMI database were used to compare two feature extraction algorithms. Contrast, dissimilarity, and asm were some of the texture features that were extracted from grayscale images using the GLCM

approach. The GLCM approach required less calculation time and was responsive to changes in picture data. Images of the brain that PD had altered were examined using the Principal Component Analysis (PCA) technique. PCA extracted eigenvectors from the co-variance matrix and transformed images into grayscale to minimize dimensionality. However, compared to GLCM, PCA required more memory and processing time. While the PCA technique only received a score of 87.5% accuracy, the GLCM approach with distance ($d=1$) and angle (0 and 90) achieved 90.5% accuracy. Results from the GLCM approach were more accurate than those from the PCA method, attaining 90.5% accuracy.

Battula Balnarsaiah and his colleague (Balnarsaiah, Nayak, Sujeetha, Babu, & Vallabhaneni, 2023) have developed a revolutionary deep learning technique Using brain MRI images for Parkinson's disease classification. Using parallel branches and extra convolutional layers, this approach modifies the ResNeXt model to improve overall performance and disease classification accuracy. The model learns patterns in people with Parkinson's disease (PD) using brain MRI scans from Kaggle. A data preparation approach is used to maintain uniformity, including data cleaning, feature scaling, and normalization. The evaluation was conducted on a dataset comprising 249 images from 2 classes for testing, and the training process utilized 582 images from the same 2 classes. The proposed method resulted in significant enhancements in both training and validation accuracy. Specifically, there was a notable reduction in training loss from 60 to 35 and a decrease in validation loss from 35 to 100%, indicative of improved model performance during the training and validation phases.

Zulqurnain Sabir and his fellow mates (Sabir, Raja, Baskonus, & Ciancio, 2023) presents a mathematical model using Levenberg-Marquardt backpropagation technique (L-MB) neural networks to study coronavirus effects, including vaccination and quarantine. The model's dynamics depend on human population subgroups, with 76% training performance, 10% testing, and 14% authorization.

Qiliang Chen and his colleagues (Chen, Sabir, Raja, Gao, & Baskonus, 2023) examine the FO-EEMM fractional order economic and environmental mathematical model, which predicts integrated economy growth in low-level economies during economic disruption. The model's structure includes control accomplishment cost, manufacturing capability, and technical exclusion diagnostics cost. The model's stochastic computing performance is evaluated using 8-neuron SCGNNS.

Kashif Nisar and his fellow mates (Nisar et al., 2021) use the Morlet wavelet activation function to build feed-forward artificial neural networks to solve pantograph differential Lane-Emden models, confirming its accuracy through stability, performance, and statistical analysis with forty trials.

Zulqurnain Sabir and his colleague (Sabir, Wahab, Javeed, & Baskonus, 2021) evaluate a stochastic numerical computing framework using Gudermannian neural networks, GA, and ASA, testing it on the higher-order nonlinear singular differential model (HO-NSDM) and comparing numerical and exact solutions, ensuring its efficiency, implication, and dependability.

Muhammad Umar and his colleague (Umar et al., 2021) proposed a study that uses Morlet wavelet artificial neural networks (ANNs) to forecast nonlinear HIV infection systems of latently infected CD4+ T cells. Sequential quadratic programming and global search of genetic algorithms are used to optimize the strategy. Statistical approximations and comparative analyses validate the method's accuracy and robustness.

Yolanda Guerrero Sanchez and her fellow mates (Guerrero Sánchez, Sabir, Günerhan, & Baskonus, 2020) discuss the importance of the stomach in maintaining human health and explore the mathematical form of the nervous stomach model. It focuses on the differential transformation method and numerical solutions, considering three compartments: tension (T), food (F), and medicine (M). The control of gastric secretion was discovered in the early 20th century.

Zulqurnain Sabir and his colleague (Sabir, Raja, Shoib, & Aguilar, 2020) present a fractional Meyer neuro-evolution-based intelligent computing solver (FMNEICS) for numerical treatment of a doubly singular multi-fractional Lane-Emden system, utilizing MWNN-GASQP, and validated through comparative studies and statistical analysis.

Zulqurnain Sabir and his mates (Sabir, Raja, Guirao & Shoib, 2021) present a new stochastic computational framework for non-linear singular fractional Lane-Emden differential equations, utilizing fractional Meyer wavelet artificial neural network (FMW-ANN) and hybrid computing's meta-heuristic power, with six scenarios tested for accuracy, convergence, stability, and robustness.

Mohamed A. Abdelkawy and his colleague (Abdelkawy, Sabir, Guirao, & Saeed, 2020) develop a second-order nonlinear Lane-Emden coupled functional differential model using the spectral collocation approach, validate it with three versions, and assess its efficacy, finding it more suitable and efficient for solving complex, single-coupled nonlinear LE second-order systems of FD equations.

Yolanda Guerrero-Sánchez and his mates (Sanchez, Umar, Sabir, Guirao, & Raja, 2018) employ feed-forward artificial neural networks optimized with particle swarm optimization and interior-point methods to solve a biological nonlinear HIV infection model. The scheme's accuracy is confirmed by comparing numerical results with Adams, and its reliability, steadiness, and precision are indicated by median and semi-interquartile range.

Rakhi Singh and her team (Singh, Mishra, & Gupta, 2023) presents a unique Caputo-Fabrizio-fractal-fractional derivative model for tumor growth, utilizing fixed-point Schauder and Banach theorems. The model is analyzed qualitatively and statistically, and numerical techniques are developed for each compartment. Matlab investigation reveals excellent results, with variations in fractional order and dimension impacting tumor growth rate, non-deviating cells, and proliferation.

Özgül İlhan and his colleagues (İlhan & Şahin, 2024) explore the use of the Morgan-Voyce Colloquiation method for predicting non-fatal illnesses in populations. It compares it to other methods like Homotopy perturbation and Laplace Adomian decomposition. The study finds the method effective in managing epidemics and calculating vaccination rates, contributing significantly to the literature.

Sanzida Solayman and her fellow mates (Solayman, Aumi, Mery, Mubassir, & Khan, 2023) developed an intelligent web application for quick diagnosis based on symptoms by automating COVID-19 identification using machine learning. The Israeli Ministry of Health's open-source dataset is preprocessed, and several classifiers are assessed. The hybrid CNN-LSTM model with SMOTE support is the best, with an accuracy of 96.34% and an F1 score of 0.98. The COVID-19 prognosis is instantly accessible to users through the application, which is hosted on a website. The study uses the LIME architecture to provide explainable AI as well. Nevertheless, drawbacks include the requirement for private datasets with more biomarkers and dependence on a region-specific dataset.

Tiwalade Modupe Usman and his fellow mates (Usman, Saheed, Ignace, & Nsang, 2023) create a reliable diagnostic model for diabetic retinopathy (DR), a major cause of vision loss globally, the study makes use of machine learning and deep learning approaches. With the use of transfer learning and pre-trained convolutional neural networks, the model is able to achieve remarkable lesion detection and classification accuracy rates of up to 94.40%. It is shown that the model can be applied in clinical settings and large-scale screening programs. To guarantee its dependability in various clinical circumstances, more testing with bigger datasets is necessary. The research emphasizes how deep learning techniques may be used to detect DR and central serous macular edema early on, which has great promise for enhancing patient outcomes and advancing public health campaigns.

Ammar Almomani and his fellow mates (Almomani et al., 2024) used a hybrid technique that combines classic machine learning classifiers with deep learning models for feature extraction, the research

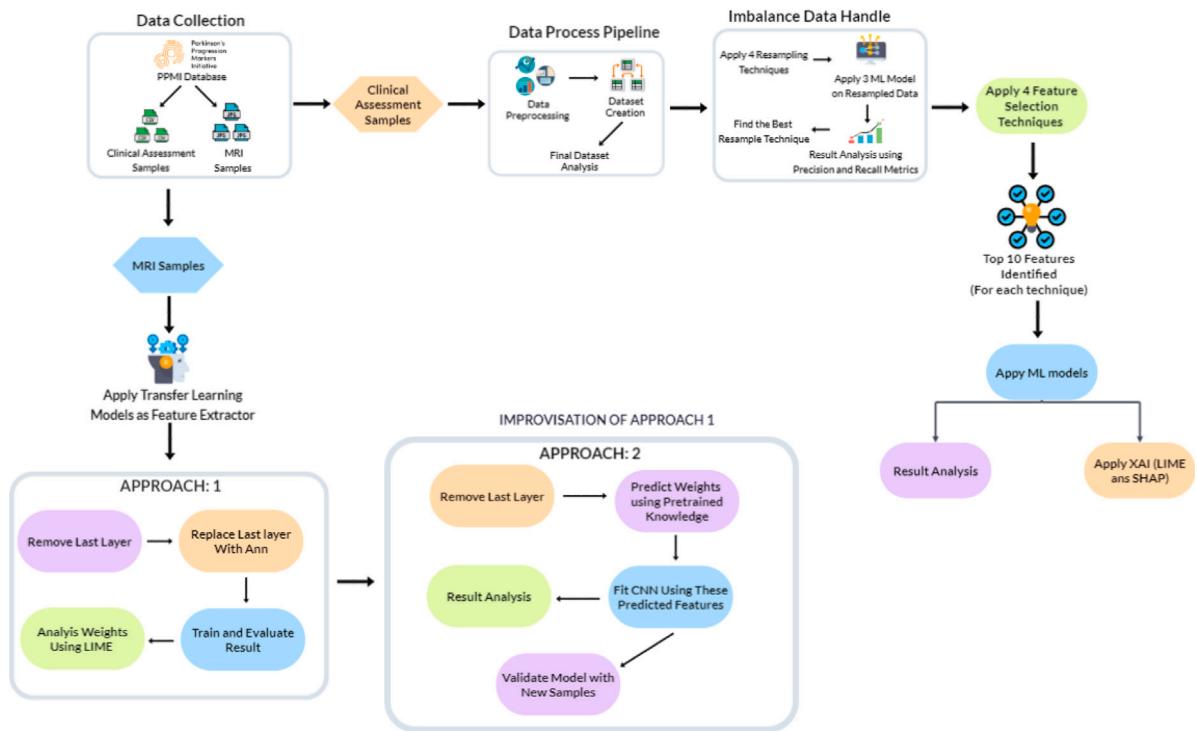


Fig. 1. Process Flowchart of our Overall Methodology.

seeks to improve the identification of cyberbullying on social media platforms. This method greatly improves understanding of intricate cyberbullying scenarios by utilizing pre-trained deep learning models such as InceptionV3, ResNet50, and VGG16. Based on a picture dataset, the hybrid strategy achieves a higher accuracy rate of 67%, demonstrating its superior performance over individual strategies. The work highlights how further developments in model architectures and deep learning approaches could lead to more accurate detection of cyberbullying. To improve detection skills even further, future plans include for gathering and categorizing additional datasets that are specific to various cultural contexts as well as adding object segmentation and recognition.

3. Methodology and model architecture

This section presents comprehensive descriptions of different components of our methodology. Fig. 1 shows an overview of the steps we applied in our study.

3.1. Data collection

This subsection provides comprehensive details about our data source and the records we collected.

3.1.1. Parkinson's Progression Marker Initiative (PPMI)

PPMI (<https://www.ppmi-info.org/study-design/study-cohorts>), a globally renowned research project to enhance knowledge of Parkinson's disease, was initiated and funded by the Michael J. Fox Foundation in 2010. Furthermore, over 30 private non-profit founders contributed funds for the establishment of this association to foster diverse cooperation among researchers, sponsors, and study participants to improve PD medications. PPMI incorporates clinical, imaging, genetic, survey, and biological samples from approximately 1500 patients worldwide to obtain a complete understanding of the condition. The ultimate objective of PPMI is to improve care and outcomes for those affected by the disorder by considerably advancing scientific knowledge of Parkinson's through collaboration and the collection of various datasets.

3.1.2. Data

In our study, two different types i.e., numerical and image samples are collected and the overall information of collected data is provided below.

1. Clinical Assessment Samples

To attain our goal, 12 sets of clinical assessment tests each with a significant number of features are collected from 2010 to 2023. These tests consist of two distinct sets of patient records: one includes individuals diagnosed with Parkinson's disease, while the other includes those without the disorder. Furthermore, these 12 Assessments explore several aspects of the patient's conditions, including motor functions activity, writing abilities, behavioral and speech patterns, sleep disturbances, and mood disorders such as depression and anxiety, alongside olfactory and autonomic sensitivities. Here, Table 1 presents the detailed breakdown of our 12 collected clinical assessment samples and features.

2. Neuroimaging Samples

In addition to the clinical assessment samples, a total of 7500 MRI samples between the years 2022 and 2023 are collected as well. These samples include both male and female patients, ranging in age from 40 to 85 years. The comprehensive breakdown of MRI samples is provided in Table 2.

3.2. PD detection using clinical assessments samples

The methodology used in the PD analysis and interpretation of the clinical assessment data is described in this part. In addition, it also provides the implementation of the Machine Learning strategy customized for this dataset. In this context, Algorithm 23 encapsulates the pseudo-algorithm encapsulating the entire process of detecting Parkinson's Disease using Clinical Assessments records, providing a comprehensive overview of the study.

Table 1

The overview of collected samples of clinical assessment.

Test	Description of functions tested	No. of records	No. of features
1	Non-motor task of daily living	17 991	15
2	Motor task of daily living	18 018	22
3	Motor function	20 718	62
4	Cognitive assessment	9968	35
5	Autonomic function	11 032	43
6	Olfactory function	665	92
7	Complications	7586	23
8	Epworth Sleepiness Scale	11 037	16
9	Geriatric Depression Scale (GDS-15)	11 354	23
10	Rapid Eye Movement Sleep Disorder	11 048	20
11	Semantic Fluency	9455	13
12	State-Trait Anxiety Inventory for Adults	11 340	46

Table 2

The overview of collected MRI samples.

Collected information	PD	Healthy
No. of samples	3500	4000
Male	40 ≤ Age ≤ 85	40 ≤ Age ≤ 80
Female	40 ≤ Age ≤ 80	45 ≤ Age ≤ 70

Table 3

The overview of the final dataset.

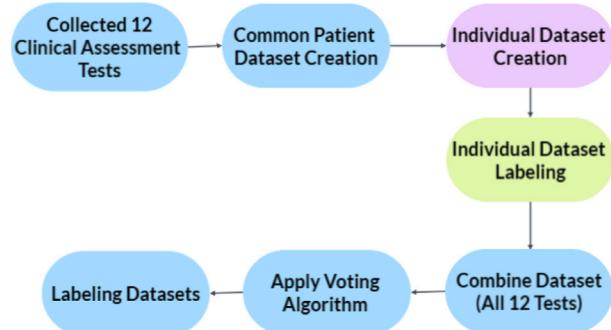
Dataset	No. of records	No. of features	PD	Healthy
Clinical Assessment	1277	273	155	1122

Algorithm 1 PD Detection Using Clinical Assessments Samples

- 1: **Dataset Preparation and Preprocessing:**
- 2: Load and preprocess the clinical assessment dataset with two classes: Parkinson's disease and healthy.
- 3: **Dataset Resampling:**
- 4: **for** each resampling technique **do**
- 5: Apply resampling technique to address data imbalance
- 6: **end for**
- 7: **Model Training with Recall Analysis:**
- 8: **for** each machine learning model **do**
- 9: Train model on resample records
- 10: Analyze precision and recall scores
- 11: **end for**
- 12: Determine the most effective resampling approach
- 13: **Feature Selection with Resample records:**
- 14: **for** each feature selection method **do**
- 15: Apply feature selection methods to find top 10 features on resample data
- 16: **end for**
- 17: **Model Training and Validation:**
- 18: **for** each machine learning algorithm **do**
- 19: Train and validate the model using the selected features
- 20: **end for**
- 21: **Explainable AI Techniques:**
- 22: Utilize LIME and SHAP on clinical assessment records
- 23: Interpret and explain the model predictions using the explainable AI techniques

3.2.1. Data process pipeline

A comprehensive overview of the Data collection, Dataset creation, Data preprocessing, and Final Dataset description procedures are given in this section.

**Fig. 2.** Process Flowchart of Dataset Creation using Clinical Assessment Tests.**1. Data Preprocessing**

To perform preprocessing on the clinical assessment dataset, a large amount of collected null and duplicate records are removed. The process of checking duplicate and null records is performed both before and after the dataset is created. However, the null entries are filled with the average and minimum values in some cases. Afterward, a few irrelevant features i.e., "Rec id", "Origin entry", and "Last update" are deleted. Finally, feature normalization is performed to ensure all numerical attributes fall into a similar range.

2. Dataset Creation

To create the dataset using these 12 clinical assessment tests, we went through multiple phases. In this research, we outline the brief idea of this dataset creation, illustrated in Fig. 2 to understand the overall idea. However, the complete procedure is elaborated in our other research titled "Single and Multimodal Analysis for Parkinson's Disease to Detect Its Underlying Factors" (Islam, Turza, Fahim, & Rahman, 2024). This research proposes to develop an AI model to analyze Single and Multimodality techniques and further identify the top importance features that lead to Parkinson's disease. The collected samples were a mix of patients who did not undergo all 12 tests. Therefore, we need to find only those patients who have taken all 11 tests. Thus, we started by creating a Common Patient Dataset by applying a search algorithm considering two parameters i.e., Patient No and Event ID. In the next phase, Individual Datasets were created by containing patients who have undergone all tests. To carry out, each of the datasets was taken into account one by one, and a search algorithm that matches up with common patient records considering two parameters i.e., Patient No and Event ID. Afterward, labels for each record of these individual datasets need to be created since the initially collected samples are unlabeled. Afterward, we created a combined dataset by merging these 12 databases. The final phase was to create labels for the combined dataset. For this, we used the common patient dataset where the Majority voting Algorithm is applied.

3. Final Dataset Description

An overview of our final dataset after preprocessing is discussed in Table 3.

3.2.2. Imbalance data handling

To resolve these issues, we have applied 4 four resampling techniques i.e., oversampling, under-sampling, and a hybrid of both techniques. Afterward, we fit and trained three distinct machine learning algorithms—SVM, AdaBoost, and GBDT, using new resample data to examine the results of two evaluation metrics: precision and recall. Since the precision analysis percentage of the model is accurately identified among all positive predictions. On the other hand, the recall analysis percentage of the model accurately identified all the actual positive

instances. Therefore, high recall indicates the accurate classification in our study. Following the recall analysis, we determined the best method to carry out our study. The four resampling techniques are as follows: (i) Synthetic Minority Over-sampling Technique (SMOTE), (ii) Adaptive Synthetic Sampling Method (ADASYN), (iii) Hybrid of SMOTE-Tomek Links Technique, (iv) Hybrid of SMOTE-ENN are used in this research are discussed below:

Synthetic Minority Oversampling Technique (SMOTE) is an Oversampling technique (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) that builds synthetic instances for underrepresented classes based on existing data points, increasing dataset balance and model performance. This eliminates biases and produces a more fair learning environment. Next, the Adaptive Synthetic Sampling Method (ADASYN) is also an Oversampling technique (He, Bai, Garcia, & Li, 2008) used in imbalanced classification tasks, particularly for underrepresented minority classes. It generates synthetic examples by adaptive adjusting the weights of minority class examples, focusing on low data density regions. This adaptability helps re-balance class distribution, producing more robust and accurate models. Afterward, the Hybrid of SMOTE-Tomek Links Technique is a combination of over and sampling techniques that aims to address class imbalance by constructing synthetic instances and refining the dataset, leading to improved model training and performance and oversampling of minority classes attached with undersampling of noisy examples by this. Lastly, the Hybrid of SMOTE-ENN Technique known as the Hybrid of Synthetic Minority Over-sampling Technique (SMOTE) and Edited Nearest Neighbor (ENN) (Yang et al., 2022) combines oversampling and under-sampling to enhance dataset quality, model generalization, and performance by creating synthetic examples for minority classes and deleting noisy.

3.2.3. Feature selection techniques

To determine the top 10 features, 4 different feature selection techniques i.e., (i) Select K Best, (ii) variance threshold, (iii) Tree-based feature selection, and (iv) Forward Feature Selection are applied to identify the top 10 features. The following provides an overview of 4 distinct feature selection methods:

Select K Best is a technique that determines the k th top features by calculating the score of each feature depending on the correlation with the target features. In this study, we adopt the Chi-Square statistical test and k as 10. chi-squared test determines if two categorical variables are significantly correlated and (1) presents the equation for the calculation.

$$X_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (1)$$

Here, O = actual count, E = expected count. Next, The Variance Threshold is a baseline approach to feature selection, removing features with a variance below a certain threshold. It removes zero-variance features, which have the same value in all samples. The threshold value is 0.7 in our case. Afterward, to conduct Tree-based feature selection, the Gradient Boosting ensemble algorithm is used to minimize bias error. It corrects mistakes made by earlier assumptions and fine-tunes its approach. Additionally, It uses mean squared error as a cost function in classification problems. Lastly, Forward Feature Selection is a technique that iteratively adds important features to a predictive model, starting with the most important one and adding more based on their performance contribution. This process continues until a stopping criterion is met or further features do not significantly improve the model's performance. In our study, we have used Sequential Feature Selector to select the best features.

3.2.4. Algorithms application

To perform multi-modality analysis, the datasets are divided into training (70%) and testing (30%) groups. Afterward, we applied 9 machine learning classifiers for each of the feature selection techniques. The ML classifiers i.e., (i) Support Vector Machine (ii) K nearest neighbor, (iii) Adaboost Classifier, (iv) Multi-layer Perceptron, and (v) Extra

Trees Classifier (Vi) Logistic Regression (Vii) Decision Tree Classifier (Viii) Random Forest (ix) Gradient Boosting Decision Tree are used to detect the individuals with Parkinson's disease. We keep the default hyperparameter for all the algorithms.

Support Vector Machine is a machine learning algorithm, used for classification and regression. The maximum separation hyperplane between several classes in the specific feature can be found using SVM algorithms. In our study, a Linear kernel is used. Random Forest is also used for classification and regression. Because of their high classification accuracy, resilience to noise and outliers, and capacity to prevent overfitting, random forests (RF) are widely used in research. AdaBoost helps to enhance the performance of the model and is called the Ensemble method. The system's capacity to weight instances based on prior classifications is responsible for its reduced overfitting, computational efficiency, and resilience to noisy and unbalanced datasets. A decision tree is a popular supervised learning method for problems with regression and classification. Logistic Regression is the statistical technique, employed to depict the correlation between a binary dependent variable and several independent variables. Extra Trees Classifier is a method that combines the results of multiple decor-related decision trees inside a forest to provide a classification result. Gradient Boosting is an ensemble algorithm used for classification and regression. To produce the final projections, one blends the forecasts from several decision trees. MLP classifier is a backpropagation-trained network structure made up of several layers of input nodes connected by a directed graph.

3.2.5. Explainable artificial intelligence application

This study adopts two different XAI approaches to interpret and explain our models. Two techniques i.e., (i) Local Interpretable Model-agnostic Explanations (LIME) and (ii) Shapely Additive explanations (SHAP) are applied to understand our testing samples. As a result, we can use these interpretable models to examine which features are more relevant for individuals with PD or healthy patients. The following discusses the two XAI models:

Local Interpretable Model-agnostic Explanations (LIME) is an AI technique that creates locally accurate and interpretable models based on individual predictions, crucial for understanding machine learning models in sensitive fields. Shapely Additive explanations (SHAP) offer insights into complex model decision-making for tabular, graphic, and text datasets. SHAP calculates Shapley values to improve the interpretability of machine learning predictions.

3.3. PD detection using neuroimaging samples

The methodology employed for the analysis and interpretation of Parkinson's Disease (PD) in the context of MRI samples is detailed in this section. Furthermore, it presents the application of a tailored Transfer Learning strategy specifically designed for this dataset. In this regard, Algorithm 22 represents the pseudo-algorithm of all Parkinson's Disease Detection Using Neuroimaging Samples to provide a comprehensive overview of the study.

3.3.1. Data process pipeline

1. Data Preprocessing

The MRI samples were initially collected in DICOM format, which cannot be used directly to train our model. Therefore, we convert these DICOM format samples to JPG format. After conversion to the JPG format, the sample dimensions turned out to be $448 \times 448 \times 3$. Therefore, Every image is resized to 224×224 pixels to obtain an accurate result. Finally, the redundant samples are filtered. Out of 7500 MRI samples, only 2500 samples were considered usable. Therefore, The remaining images are excluded due to their vague quality, and no notable features are identified that could be potentially extracted for further analysis. Fig. 3 displays two of those samples that lack any distinguishing characteristics.

Algorithm 2 PD Detection Using Neuroimaging Samples

```

1: Dataset Preparation and Preprocessing:
2: Load and preprocess the image dataset with two classes: Parkinson's
   disease and healthy.
3: Approach 1: Transfer Learning with ANN:
4: for each pretrained_model in {list of selected pretrained models}
   do
5:   Load pretrained_model
6:   Replace the last layer with an ANN for binary classification
7:   Train and test the modified model on preprocessed_data
8:   Analyze important weights using LIME
9: end for
10: Approach 2: Weight Prediction:
11: for each pretrained_model in {list of selected pretrained models}
    do
12:   Load pretrained_model
13:   Remove the Last layer
14:   Extract weights of MRI samples using knowledge of the
      pretrained model and store them
15:   Add Custom CNN
16: end for
17: Model Training and Testing:
18: Train and test a CNN with the predicted weights
19: Comparative Analysis and Results Evaluation:
20: Compare results from Approach 1 and Approach 2
21: Compare results from CNN
22: Evaluate and discuss the performance of each approach

```

Table 4

The overview of the final dataset.

Dataset	No. of records	PD	Healthy
Neuroimaging	2500	1236	1244

Table 5

The parameters of the artificial neural network model.

Parameter	Value
Hidden layers	2
Neurons on each hidden layer	30 (first) 40 (Second)
Output layer	1
Activation function	Tanh
Optimization method	Adam
Loss function	Binary cross-entropy
Epochs	10
Batch size	50

2. Dataset Creation

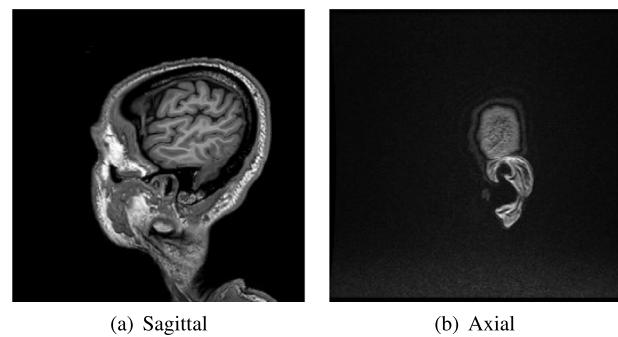
To create an MRI dataset, 3D-T1-weighted images were chosen for both PD and Healthy samples. Furthermore. Each of the samples is analyzed and chosen for the final dataset. Out of all angles, Axial and Sagittal MRI samples are chosen for our final data set, as shown in Fig. 4.

3. Final Dataset Description An overview of our final dataset after preprocessing is discussed in Table 4.

In this study, 4 different pretrained models i.e., (i) DenseNet169, (ii) VGG16 and VGG19, (iii) Resnet, and (iv) InceptionV3 from Keras are applied to analyze MRI samples. These pretrained models are already trained with a large dataset named Imagenet. Furthermore, to detect PD Using MRI samples, we are using these pretrained models to extract features to create an automated model to analyze this neurological disease deeply.

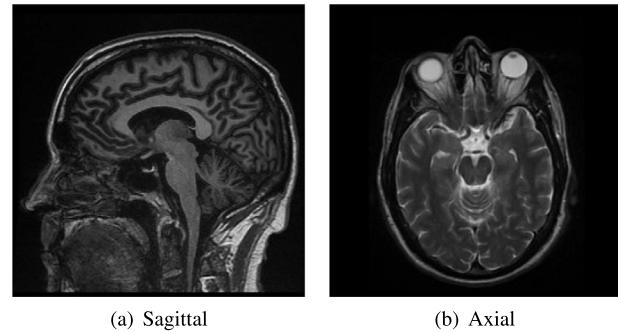
3.3.2. APPROACH 1: Pretrained models as feature extractor

The pretrained models consist of sequential layers, including convolutional, fully connected, and output layers. For the first Approach,



(a) Sagittal (b) Axial

Fig. 3. Redundant Samples.



(a) Sagittal (b) Axial

Fig. 4. The MRI Samples for Final Dataset.

we only utilize the convolutional layer while replacing the remaining two layers with our own Artificial Neural Network (ANN). Afterward, we freeze the conventional layers (which means this layer will no longer be trained, only use its pre-trained weights for our training), and using those pre-trained weights, we train and evaluate our MRI samples with our custom ANN. Following the training phase, we examine the predictive weights of MRI samples using the LIME explainable AI model. The ANN model parameters utilized in this study are shown in Table 5.

The formulas for computing *Tanh* activation function are displayed in (2).

$$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2)$$

Fig. 5 presents the key outline of Approach 1 for better understanding. Here, the pretrained layers of the VGG16 model are shown. The last layer of this model is removed and our own customized ANN is added to train with our collected MRI images. This is how we implement Approach 1, using all of the pretrained models.

The 4 transfer learning Pretrained architectures that are used in this study are discussed below:

The Densenet169 architecture is composed of multiple dense blocks, including 169 layers, 3 transition layers, 1 classification layer, and 1 convolutions layer. Each block is connected in a feed-forward style, with convolution and pooling occurring between them. The transition layers consist of a batch normalization layer, a convolutions layer, and an average pooling layer. Next, VGG16 and VGG19 are image recognition models with similar architectures, using 16 weight layers and convolutional blocks, and 19 weight layers, respectively. VGG16 incorporates dense layers for ImageNet classification, while VGG19 goes deeper into feature extraction. Both models have 64 filters and convolutional blocks, VGG16 integrates two dense layers with 4096 units, ReLU activation, and dropout for regularization, and lastly, a final dense layer with 1000 units for ImageNet classification. Afterward, ResNet-50 uses a 7×7 convolutional operation using 64 filters, batch normalization, ReLU activation, and max-pooling to reduce spatial

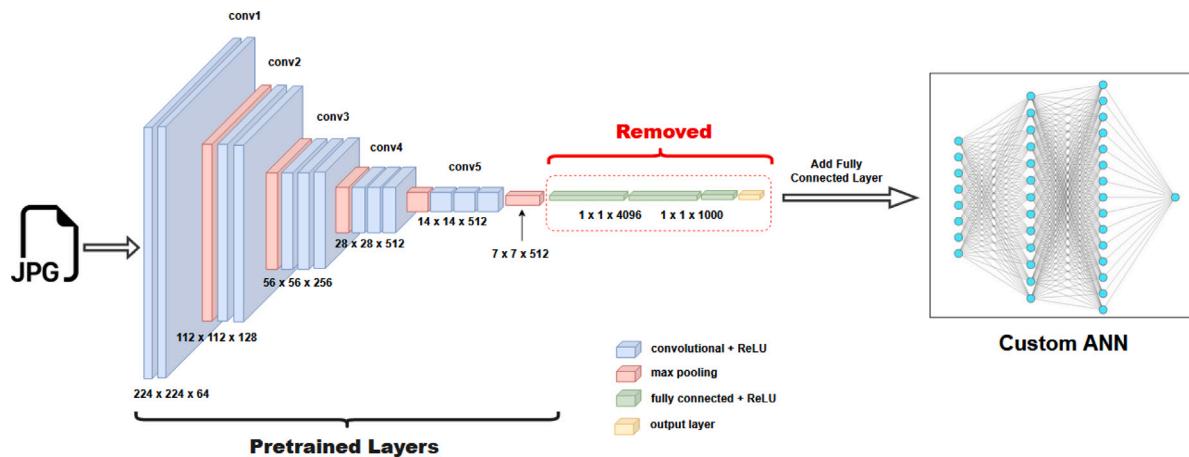


Fig. 5. The Overall Technique of APPROACH 1.

Table 6
The parameters of the convolutional neural network model.

Parameter	Value
Initial learning rate	0.001
Activation function	Relu & Softmax
Optimization method	Adam
Loss function	Binary cross-entropy
Epochs	10
Batch size	50

dimensions. It features residual blocks with four types and first uses a dense layer, and then a global average pooling layer for ImageNet classification. Finally, InceptionV3 uses factorization techniques in its modules, preferring two consecutive 3×3 convolutional layers over a single 5×5 layer. It avoids fully connected layers and integrates global average pooling, reducing parameter count and countering overfitting.

3.3.3. APPROACH 2: Improvisation of feature extraction technique

To achieve better performance, we improvise the previous approach to extract weights. The previous technique was also applied here, however, without replacing the last layer directly, we predict the weights of our MRI samples using the knowledge of the pretrained model and stored them. Simply put, The pretrained models are essentially being used as feature extractors, and the features are captured by the convolutional and pooling layers of VGG16. Notably, the obtained features are the result of the activations from the last layer before fully connected layers. Afterward, we train our own CNN model using these stored weights and evaluate the performance. Finally, validate the model was validated using the new samples. Table 6 specified the CNN model parameters used in this study.

For clear understanding, Fig. 6 is presented which outlines the overall idea of Approach 2. Here, the pretrained layers of the VGG16 model are shown. The last layer of this model is removed in the same manner as approach 1. Afterward, The features are extracted by the convolutional and pooling layers of VGG16 and our own customized CNN is trained using these extracted features with our collected MRI images.

4. Result analysis

This section provides an in-depth overview of the experiment and the outcomes that are obtained through the methodology we adopt for this study.

4.1. Performance evaluation of the PD detection of clinical assessments samples

This section deeply analyzes the performances of different approaches using clinical assessment samples.

Following the partition of the dataset into a 7:3 ratio, our model underwent training with 70% of the records. Upon completion of the training phase, the remaining 30% of the records are employed to assess the model's performance. Hence, the dataset was divided into 893 records for training and 384 records for testing. Additionally, a comprehensive analysis of the overall model performance is conducted using various evaluation metrics such as Accuracy, Precision, Recall, and F1-Score, utilizing the entire dataset. This thorough evaluation provides a comprehensive understanding of the model's efficacy in handling diverse data points and making accurate predictions. The formulas for computing these metrics are displayed in (3), (4), (5), and (6).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1 - Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

Where, TP and TN indicate true positives and true negatives, respectively, whereas FP and FN indicate false positives and false negatives.

4.1.1. Find best resampling technique

To determine the optimal Resampling method, 4 distinct sampling techniques are utilized. After that, we use 3 distinct machine learning models to train these fresh resample records. Next, we examine each model's precision and recall performance. Hence, the resampling technique that achieves the highest performance is recognized as the Best approach for handling imbalanced data. Table 7 demonstrates the records information of before and after application of the resampling technique.

The effectiveness of different algorithms using the resampling method is displayed in Table 8 where the values show only in Precision and Recall. Here several methods are used which are SMOTE, ADASYN, SMOTE-TOMEK, and SMOTE-ENN. Analyzing all of these methods we can see in Recall of SMOTE-ENN perform better on SVM, ADABoost, and GBDT respectively 81.25%, 89.58%, 81.25%. Similar to Precision, the SMOTE-TOMEK was 62.26%, 80.64% higher than that of the other four methods. Additionally, in Precision, SMOTE-ENN showed the lowest Precision (46.98%, 51.80%, and 53.42%, respectively).

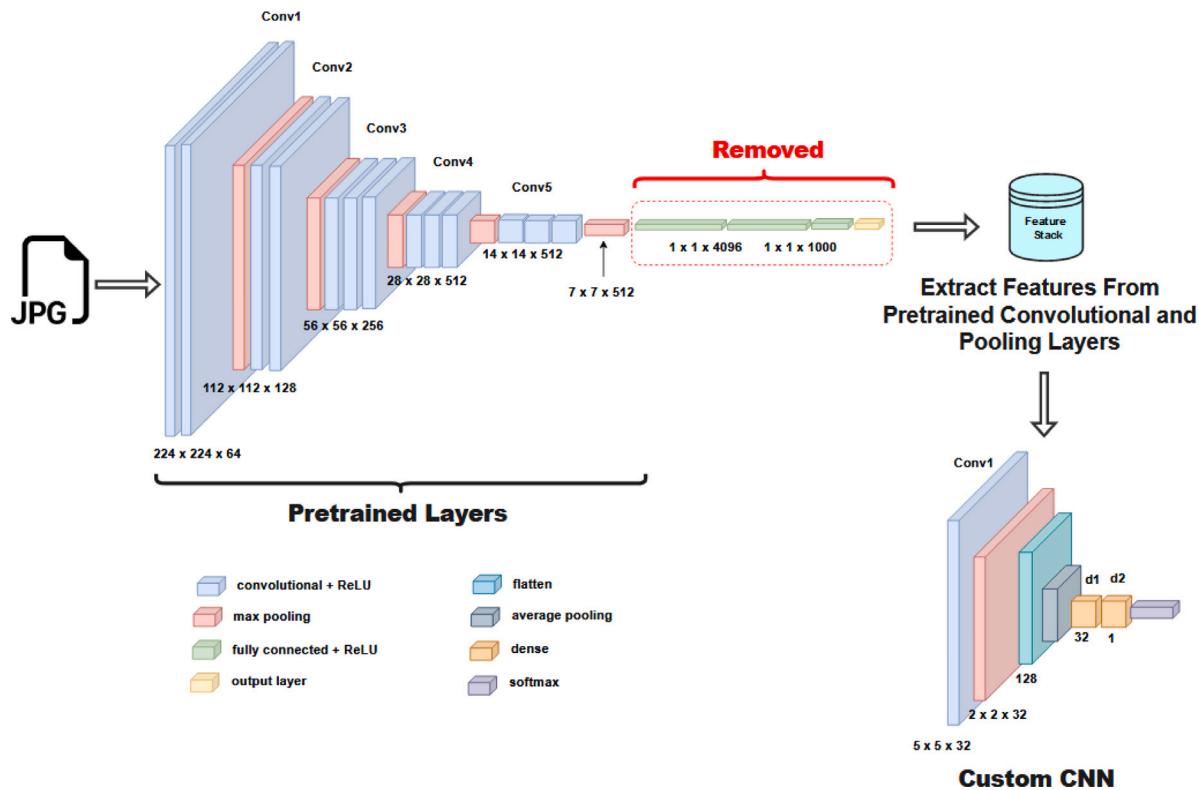


Fig. 6. The Overall Technique of APPROACH 2.

Table 7

Total number of records before and after resampling.

Sampling method	PD	Healthy
Before resampling	114	779
SMOTE	779	779
ADASYN	745	779
SMOTE & Tomek Link	779	779
SMOTE & ENN	779	482

Table 8

Performance metrics for different algorithms after using sampling and nonsampling method.

Sampling method	Evaluation metrics	SVM	ADAboost	GBDT
Imbalance Data	Precision	62	73.68	66.66
	Recall	64.58	58.33	58.33
Smote	Precision	61.53	79.41	77.14
	Recall	66.66	56.25	56.25
Adasyn	Precision	62.26	74.28	70
	Recall	68.75	54.16	58.33
SMOTE-Tomek	Precision	62.26	80.64	66.66
	Recall	68.75	52.08	50
SMOTE-ENN	Precision	46.98	51.80	53.42
	Recall	81.25	89.58	81.25

We can see that all four Resampling Techniques in Fig. 7 achieve results in SVM, AdaBoost, and GBDT. In SMOTE SVM gives the highest value which is 66.66% and GBDT shows the lowest value, 56.25%. According to ADASYN again SVM gives the highest value at 68.75%. Analyzing the hybrid resampling technique SMOTE-Tomek link, we can see SVM again shows the highest value of 68.75% and AdaBoost gives 52.08%, GBDT shows 50% value. IN SMOTE-ENN hybrid Technique AdaBoost performs best which is 89.58% but GBDT and SVM perform poorly with a value of 81.25% in both. Therefore, we can say that, smote-ENN is the best technique for data imbalance handling.

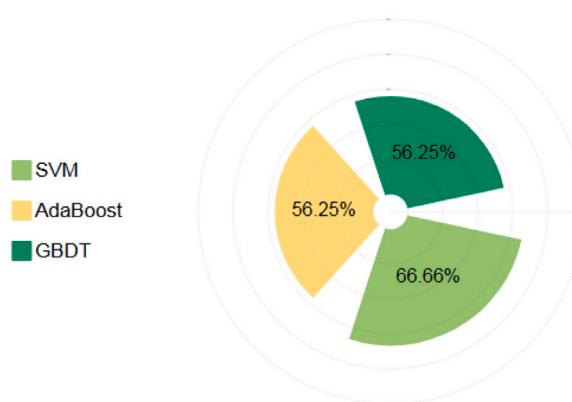
4.1.2. Identifying top features

Fig. 8 illustrates the top 10 features determined by Select K Based, Variance Threshold, Tree-Based, and Forward Feature selection are T1 _ NUPSOUC, T6 SCENT _ 40 _ RESPONSE, T6 SCENT _ 40 _ CORRECT, T6 SCENT _ 38 _ RESPONSE, and T1 NP1COG etc consistently outperforms all other tests, and ranking among the top 10 features. Test 6 performs better almost every single time with the most essential features. Additionally, Test2, Test17, and Test4 perform better than others, and considering higher importance value, the features i.e., NP2HYGN, MCAYR, and STAIAAD are the most important found from these tests. Noted that Variance Threshold only assigns Binary value (True = 1, False = 0) while selecting important features.

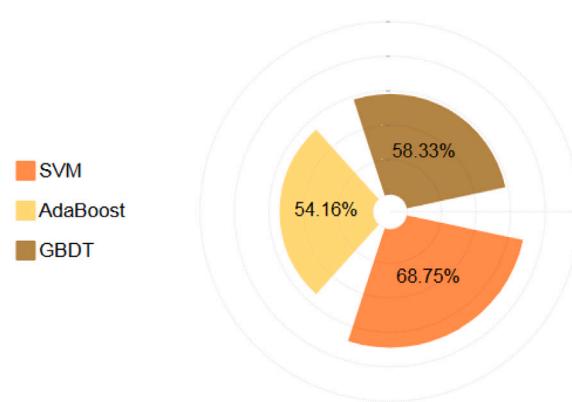
4.1.3. Performance analysis using top features

In Table 9 Select K Best analysis, the study looks at how machine learning algorithms perform. It finds that Extra Trees (ET) regularly beats GBDT in terms of accuracy, recall, and F1 score, while GBDT performs better in precision. At 96.11%, ET has the highest accuracy level, followed by 95.71% for GBDT and RF. Additionally, KNN outperforms Adaboost, with KNN demonstrating 100% precision. ET's F1 score is the highest at 94.95%. Similarly, RT outperforms RF and GBDT in precision (90.72%), but RT outperforms them in accuracy, recall, and F1 score (92.61%). In all sectors, LR performs the worst. Additionally, the study examines how well-tree-based methods perform when employing the Adaboost, GBDT, ET, KNN, MLP, RF, LR, DT, and SVM (Linear) algorithms. With 98.44% accuracy, ET had the highest level of performance, while KNN, GBDT, and RF achieved similar levels (97.66%), (97.28%), and (97.28%). KNN had the highest recall (100%) and GBDT had the highest precision (98.08%). On the other hand, if we utilize the forward feature selection method, ET shows the highest accuracy which is 92.61%, and the same things happen in precision, recall, and F1-Score. Logistic Regression performs worst in every metric. Finally, we can say Select K Best analysis reveals that Extra Trees (ET) consistently outperforms GBDT in accuracy, recall, and F1 score, while GBDT excels in precision. KNN outperforms Adaboost and RT, while Logistic Regression performs worst in all metrics.

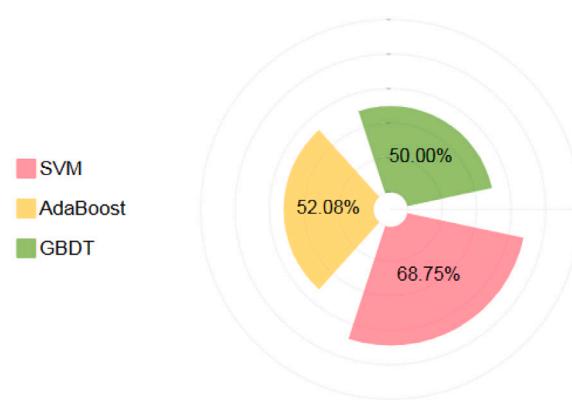
Resampling Technique: SMOTE



Resampling Technique: ADASYN



Resampling Technique: SMOTE-Tomek Links



Resampling Technique: SMOTE-ENN

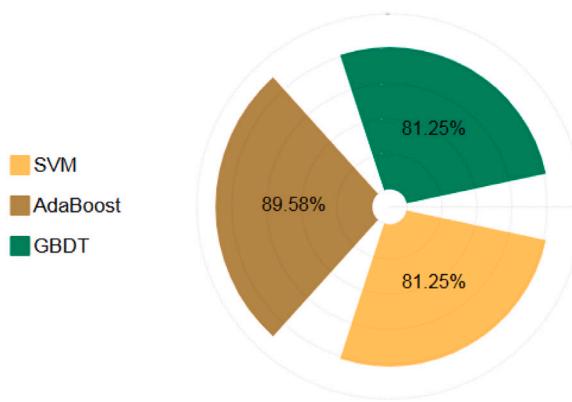


Fig. 7. Recall Analysis of Different Model Using Different Resampling Methods.

Table 9

Performance analysis of different models using different feature selection method.

Feature selection methods	Evaluation metrics	SVM	RF	LR	DT	KNN	MLP	Adaboost	GBDT	ET
Select K Best	Accuracy	91.38	95.71	91.05	92.99	95.33	93.38	95.33	95.71	96.11
	Precision	86.27	93.14	86.27	93.14	88.23	87.25	93.14	94.12	92.16
	Recall	92.63	95.96	90.72	89.62	100	95.70	95	95.05	97.92
	F1-Score	89.34	94.52	88.44	91.35	93.75	91.28	94.05	94.58	94.95
	Accuracy	76.65	92.22	76.65	87.54	89.88	84.43	87.94	89.88	92.61
Variance Threshold	Precision	77.31	90.72	67.01	83.50	83.50	78.35	87.63	87.63	89.69
	Recall	66.37	88.88	69.89	83.50	89.01	80	81.73	85.86	90.62
	F1-Score	71.43	89.79	68.42	83.50	86.17	79.17	84.58	86.73	90.15
	Accuracy	95.33	97.28	94.55	92.61	97.66	96.50	96.89	97.28	98.44
Tree Based Method	Precision	90.38	98.08	89.42	93.34	94.28	94.23	97.11	98.08	97.11
	Recall	97.91	95.32	96.87	90.48	100	97.03	95.28	95.33	99.02
	F1-Score	94	96.68	93	90.90	97.02	95.61	96.19	96.68	98.06
	Accuracy	76.65	90.66	75.10	89.10	85.60	82.49	91.05	92.21	92.61
Forward Feature Selection Method	Precision	77.65	93.61	77.47	91.49	90.42	81.91	93.64	95.74	96.81
	Recall	65.18	83.01	63.64	81.13	75.22	73.33	83.81	84.90	85.05
	F1-Score	70.87	88	68.63	86	82.12	77.39	88.44	90	90.55

4.1.4. Performance analysis explainable AI

In this study, we use the LIME technique to interpret and visualize our model more clearly. Fig. 9 demonstrates the Interpretation of a PD and health case prediction using LIME explainable AI. To analyze, we took 2 different samples from the test dataset and applied ET. The lime model predicts this sample as PD with the 64% percentages. Furthermore, the features that are marked as “orange color” lead this sample to be predicted as PD. On the other hand, the other sample is predicted as Healthy with 96% percentages where, except the single features, all are voting for the Healthy case.

Fig. 10 demonstrates the Interpretation of feature correlation for a PD and health case prediction using LIME explainable AI. Here, the green and red colors resemble of PD and healthy class, respectively. Here, Fig. 10 shows the red marked features are highly correlated to class 0 which means the healthy class, whereas, the green marked a healthy class with significant values. Conversely, from Fig. 10, only a single feature is correlated to class PD, and the rest are to class Healthy with their significant value.

Another XAI technique is SHAP, which we used in this study. To apply SHAP, we have applied the ET model and determined the Shapley value to explain the top 10 features.

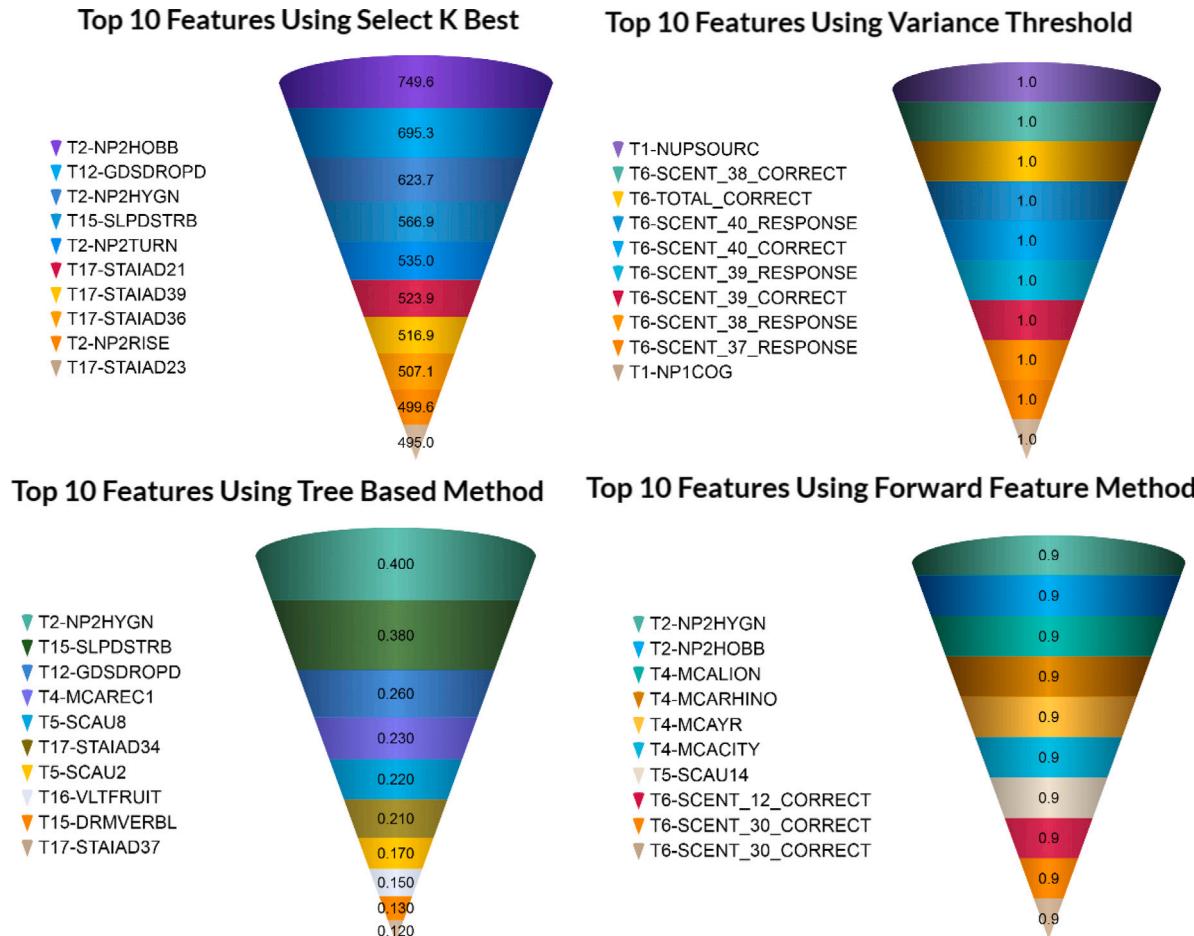


Fig. 8. Top 10 Important Features for Different Feature Selection Techniques.



Fig. 9. Interpretation of a PD (left) and Healthy (right) Case Prediction using LIME Explainable AI.

Fig. 11 illustrates the values of the top 10 features for the prediction of the healthy samples. This method makes it easy to understand how each feature contributes to the model's output. Here, the base value is 0.5, and the blue and red marked features contribute to driving below and above the base value, respectively. Therefore, blue and red marked features are class healthy and PD, respectively. From Fig. 11, we can see there is a base value is 0.5 and a predicted value of 0.02. We can also see the T4-MCAREC-2 feature is driving above the predicted value. On the other hand, T15-SLPDSTRB and T12-GDSDROPD are driving below the predicted value. In the graph, the direction shows

the importance of features driving the above or below the predicted value. This visualization also helps to clarify which features are offset.

Fig. 12 demonstrates how each feature is impacting the target class and to what extent. This SHAP explainer creates a visual representation for all samples in our dataset using Shapley values grouped by features on the y-axis. Additionally, the features are listed in descending order of their importance. Besides, for each group, the color of the pointer determines the importance value of the same features; that is, high-importance values for the features are red, and low-importance values are blue. Furthermore, samples that are on the right and left sides of the Shapley value 0, have a positive and negative impact on prediction,

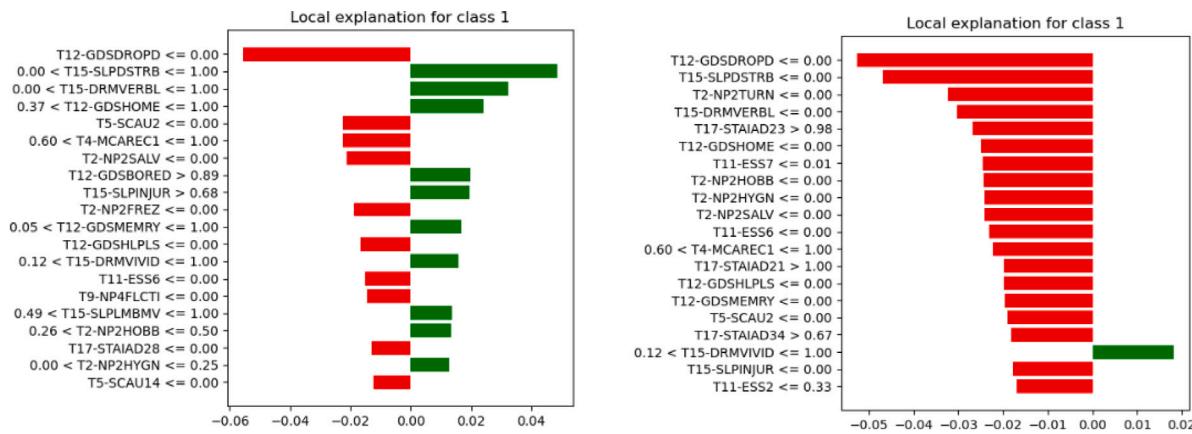


Fig. 10. Interpretation of Feature Correlation to Class using LIME Explainable for PD (left below) and Healthy (right below) case.



Fig. 11. Interpretation of features importance using SHAP explainable.

Table 10
Performance analysis of ANN using different feature extraction techniques (Approach 1) (%)

Architectures	Train accuracy	Train loss	Test accuracy	Test loss
DenseNet169	48.92	0.78	34.76	0.77
VGG16	48.92	0.74	49.76	0.75
VGG19	49.02	0.73	48.24	0.75
Resnet50	48.92	0.76	45.11	0.76
InceptionV3	51.08	0.76	50.24	0.76

respectively. The greater the distance from 0 is, the greater the positive or negative impact on the target class. Therefore, from Fig. 12, it can be seen that the first feature is T12-GDSDROPD, the red dots are residing on the right side of the Shapley values, thus, high-importance values have a positive impact on the prediction. The third feature is 15-SLPDSTRB, where the blue dots lie on the right side of the Shapley value, that is, lower importance values of these features have a positive impact on the prediction, while high values have a negative impact.

4.2. Performance evaluation of the PD detection of neuroimaging samples

This section discusses the in-depth performance of transfer learning models. To accomplish this, we split the entire image dataset into a 7:3 ratio. Consequently, 70% of the total images are employed to train our transfer learning model, while the remaining 30% are reserved for testing purposes. Hence, the dataset was divided into 750 MRI samples for training and 1750 samples for testing. In addition to that, the overall model performance is analyzed utilizing the accuracy and loss matrices.

4.2.1. Performance analysis APPROACH 1

In the provided Table 10, VGG16 achieved the highest test accuracy at 49.76%, while ResNet50 and VGG19 both demonstrated the second-highest accuracy. Surprisingly, DenseNet169 performed poorly, also yielding a test accuracy of 49.76%.

According to Table 10, VGG16 stands out with the highest accuracy of 49.76%. To gain a deeper understanding of VGG16's performance, Fig. 13 is generated to visualize training and validation accuracy as

well as loss throughout 10 epochs. The graph we are looking at shows a consistent, linear trend. This implies that accuracy and loss remain unchanged as we proceed through each epoch. Since this line does not go up, it tells us that our model is not improving its accuracy or reducing its mistakes as we keep training it. This lack of improvement suggests the need for a closer examination of the model's weights to identify potential adjustments that could enhance overall performance and yield more accurate results.

4.2.2. Analysis of important weights using LIME

To analyze the important weights of our MRI images, we utilize the Lime Interpretation model. Here, we take a PD case to interpret features. We analyze all transfer learning models with LIME, to interpret the important weights.

Fig. 14 illustrates the LIME interpretation of the MRI sample after the application of Approach 1. To interpret the sample, LIME explainer is utilized to plot 1000 top superpixels which creates the highest positive influence on a particular class prediction. Afterward, the explanation is plotted onto a heatmap visualization for better understanding. According to Fig. 14, the color bar of the heatmap indicates the values of weights. Here, 0 is the threshold value, greater than zero that is 'blue region' represents the positive towards PD Class, and less than zero that is 'red region' represents the positive weight towards the Healthy class. For a clear idea, consider Fig. 14(b), for in-depth discussion. There are multiple dark blue regions can be seen in Fig. 14(b). Therefore, the weights of these regions have the highest importance for predicting this sample as PD Class. Additionally, the weights of lighter blue color regions show the least importance for PD Class prediction. On the flip side, the weights of dark red color regions lead more towards to healthy class. Thus, deeper color intensity represents the higher importance weights for Corresponding classes.

For in-depth understanding, we take a sample that fits the best model in approach 1 and apply LIME explainer.

Fig. 15(a) illustrates that 1000 superpixels are most positive towards the PD Class while hiding the rest of the image. Simply put, the region is mainly contributing Mostly for predicting this sample as PD Class. Fig. 15 (b) shows the same sample, however, the rest of the image is present in this case. Therefore, the difference between the overall

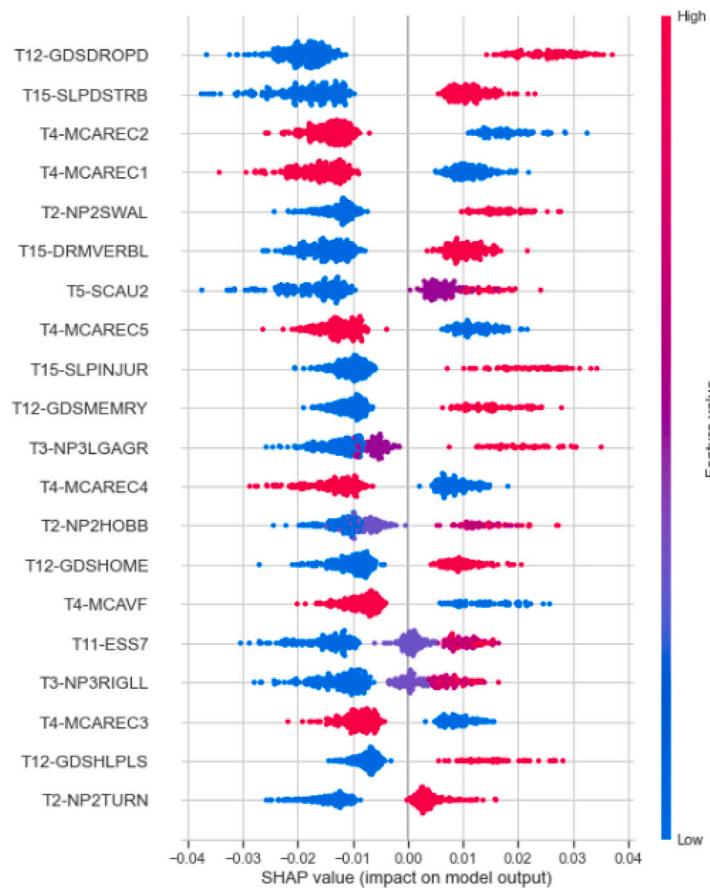


Fig. 12. Interpretation of Features for Shapley Value using SHAP.

image and the positive region can clearly be understood from this visualization Lastly, Fig. 15(c) shows the superpixels with a negative impact on the PD class, while hiding the rest of the image. In other words, these weights predict this sample as a Healthy class. Lastly, Fig. 15(d) is the sample containing the whole image.

In order to improve results, our study adopts another new method, by improvising APPROACH 1. In this new approach, we only pay attention to the important weights shown in Fig. 14. The goal of this deliberate focus on particular weights is to maximize efficiency and get better outcomes. Our goal in focusing on these crucial characteristics is to optimize the model for increased precision and effectiveness. This method enables a focused refinement process by maximizing the performance of the model overall by utilizing the importance of particular weights.

4.2.3. Performance analysis APPROACH 2

From Table 11 description, We have analyzed the performance of CNN using different feature extraction techniques. In training accuracy, we achieved the highest result in Resnet50 which is 83.57%, and the lowest result in InceptionV3 which is 16.94%. Here we see that DenseNet169 has a low train loss which is 0.42. we can say for train DenseNet169 got the best result. In test accuracy, we get the highest validation accuracy in DenseNet169 which is 85.08%. But we got the lowest validation loss in InceptionV3 which is 0.50. Since our test accuracy for DenseNet169 is higher than the training accuracy, it means our model has trained well.

In this paper, we have used six models, such as Only CNN, DenseNet169, VGG16, VGG19, Resnet50, and InceptionV3, and we did 10 epochs. Fig. 16(a) is only the CNN model. In this graph, we see that for the 10 epoch, the training and validation accuracy is the same. Fig. 16(b) is for the DenseNet169 model. It demonstrates that until the

Table 11
Performance analysis of CNN using different feature extraction techniques (Approach 2) (%)

Architectures	Train accuracy	Train loss	Test accuracy	Test loss
Only CNN	65.65	3.47	66.13	0.51
DenseNet169	62.10	0.42	85.08	0.44
VGG16	49.60	0.21	34.07	0.38
VGG19	48.54	0.31	64.92	0.38
Resnet50	83.57	0.62	67.74	0.70
InceptionV3	16.94	0.43	30.34	0.50

second epoch, training accuracy rises; beyond that, it falls. However, validation accuracy fluctuates; it rises for two epochs before falling, but then rises again in three epochs and then decreases once again in four epochs. Fig. 16(c) is for the VGG16 model. In this Figure, we can see that training and validation accuracy goes down by every epoch. Fig. 16(d) is for the VGG19 model. In Figure, we see that at the 2nd epoch, the validation accuracy goes very high. The training accuracy also performs a little bit the same but after the 5th epoch, it goes parallel. Fig. 16(e) is for the Resnet50 model. This figure shows unusual performance. The training accuracy gives a stable result, but the validation could be more steady. Fig. 16(d) is for the InceptionV3 model. In this graph, the validation accuracy goes higher in 2nd epoch but on the other hand, training accuracy decreases at that moment. rest of the situation both training and validation accuracy perform almost the same.

In Fig. 17, we see the loss analysis of different architectures for training and testing. Fig. 17(a) is only trained with the CNN model. It shows that, at 1st epoch, the training loss is below 0.3, but in the next epoch, the loss is 0.4. The validation loss at the first epoch is below



Fig. 13. Performance Analysis of VGG16 in Approach 1.

0.5 and the next epoch is below 0.1. Fig. 17(b) is for the DenseNet169 model. In Figure (b), the training loss is above 0.80, but with this architecture, there is less validation loss. Fig. 17(c) is for the VGG16 model. It shows that training loss is higher than validation loss. On the second epoch, the validation loss is below 0.2, on the other hand, the training loss is 0.6. The Fig. 17(d) is for the VGG19 model. In Figure (d), we see that at the 2nd epoch, the validation loss is below 0.1, but it goes too high again. The training loss is above 0.4 at 2nd epoch, but in the next epochs, it stays around 0.58 to 0.61. Fig. 17(e) is for the Resnet50 model. It shows unusual performance. The validation loss decreases by every epoch but the training loss performs the opposite. Fig. 17(d) is for the InceptionV3 model. In this Fig. 17(d), the training loss is above 1.0 at the 1st epoch, but after that, it comes to below 0.7 at the 2nd epoch. on the other hand, validation loss is 0.65 at the 1st epoch, and the rest of the epoch both perform almost the same. In this figure, the validation loss is very high, above 1.4 in every epoch.

4.2.4. Model validation

In this study, we have created a foundational framework, essentially a blueprint, that can be deployed as a user-friendly interface or application, which is depicted in Fig. 18. To ensure the accuracy of our

model, we embarked on a validation process using unseen samples. This validation involves testing the model with new samples that are not part of our initial dataset. It is important to note that these samples were collected in October 2023, after we completed the training and testing phases with our original dataset.

In the validation phase, each of these new images is sequentially input into the model to assess its identification accuracy. Fig. 18 showcases the outcomes generated by our most effective model, specifically DenseNet169. For instance, the first sample was originally identified as a Parkinson's Disease (PD) sample, and our model correctly categorizes it as such. Similarly, the third sample, initially collected as a Healthy sample, is accurately identified as such by our model. Additionally, the rest images are PD samples, which are also identified correctly by our model.

4.2.5. Comparing our research with previous studies

Previous research in this field traditionally focused on a singular dataset. In contrast, our study uniquely explores a variety of data types, including both clinical assessment records (numerical samples) and MRI samples. In contrast to others, our studies center around both

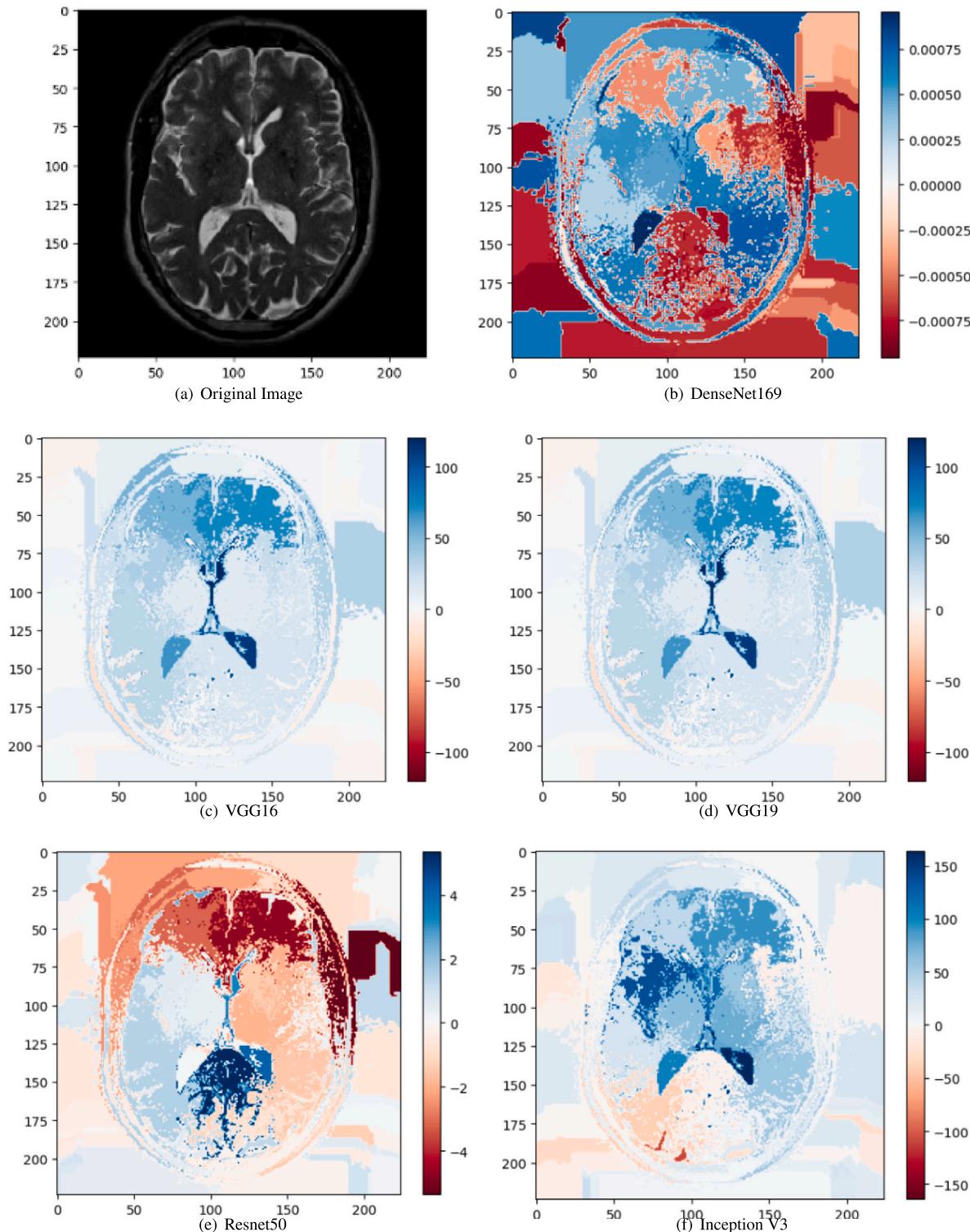


Fig. 14. Interpretation of Important Features weights of Different Transfer Learning models using LIME.

machine and transfer learning techniques. Additionally, our research aims to uncover the underlying factors using various feature selection techniques, setting our approach apart from previous work in this area. Moreover, during the examination of clinical assessments, our research implemented four distinct resampling techniques and conducted a comprehensive analysis of data handling methods. Notably,

our studies feature the prominent use of Explainable AI, with the application of both LIME and SHAP to interpret clinical assessment records. Additionally, Lime is utilized to comprehend and interpret the crucial predicted weights associated with MRI samples, contributing to our understanding of the transparency of the models. A summary

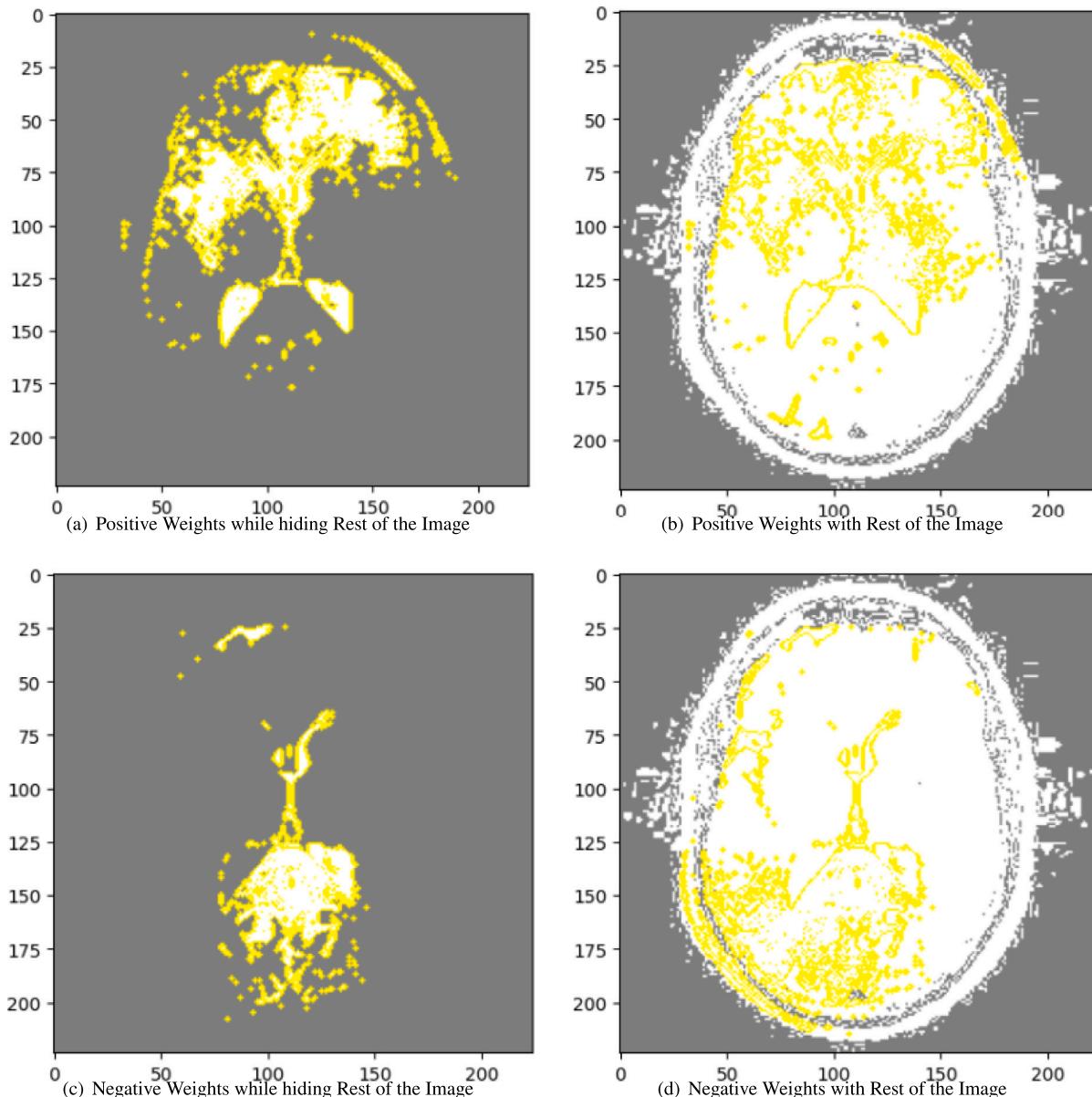


Fig. 15. Interpretation of Positive and Negative impacting Features Weights of Densenet169 for PD case.

of relevant comparative studies, including our work, is presented in Table 12.

5. Discussion

This study aimed to investigate the contributing variables to Parkinson's disease development and create machine learning and transfer learning methodologies for the disease's identification. Our goal was to find the underlying causes of the disease by doing several analyses. As per studies (Armstrong & Okun, 2020a, 2020b; Connolly & Lang, 2014), clinical experts have identified prevalent and early symptoms, including tremors, bradykinesia (slowness of movement), and rigidity (stiffness of muscles), collectively referred to as the “cardinal features” for diagnosis. These distinctive symptoms are commonly observed in the majority of Parkinson's patients during various medical assessments. In addition to these primary features, other significant aspects such as synchronizing eye movements and controlling their motion, depression, anxiety, fatigue, and sleep disorders manifest in a substantial proportion of patients, providing a comprehensive understanding of the

multifaceted nature of Parkinson's disease. In our study, we employed 12 different clinical assessment tests, identifying essential characteristics that align with the observations documented in the literature mentioned above here by medical professionals. The key features identified in our study included T1_NUPSOURC, T1_NP1COG, T2NP2HYGN, T4MCAYR, T4MCAREC1, T6_SCENT_40_CORRECT, T6_SCENT_38, T6_SCENT_40_RESPONSE, T15SLPDSTRB, T17STAID, among others. These features originate from various assessments, including non-motor and motor tasks of daily living, cognitive assessments, rapid eye movement sleep disorder, and the State-Trait Anxiety Inventory for Adults. Our findings indicate that individuals exhibiting anxiety disorders, posture difficulties, movement issues, muscle weakness, and olfactory and eye movement challenges may face an increased risk of developing Parkinson's disease. The study uses four methods to analyze clinical assessment records, identifying a hybrid of SMOTE and ENN as the best method for handling imbalanced data. Four feature selection methods are used to identify the top 10 features, and nine machine learning algorithms are used to train and validate the model. Explainable AI techniques like LIME and SHAP are used for interpretation. Apart from that, the study uses four pre-trained

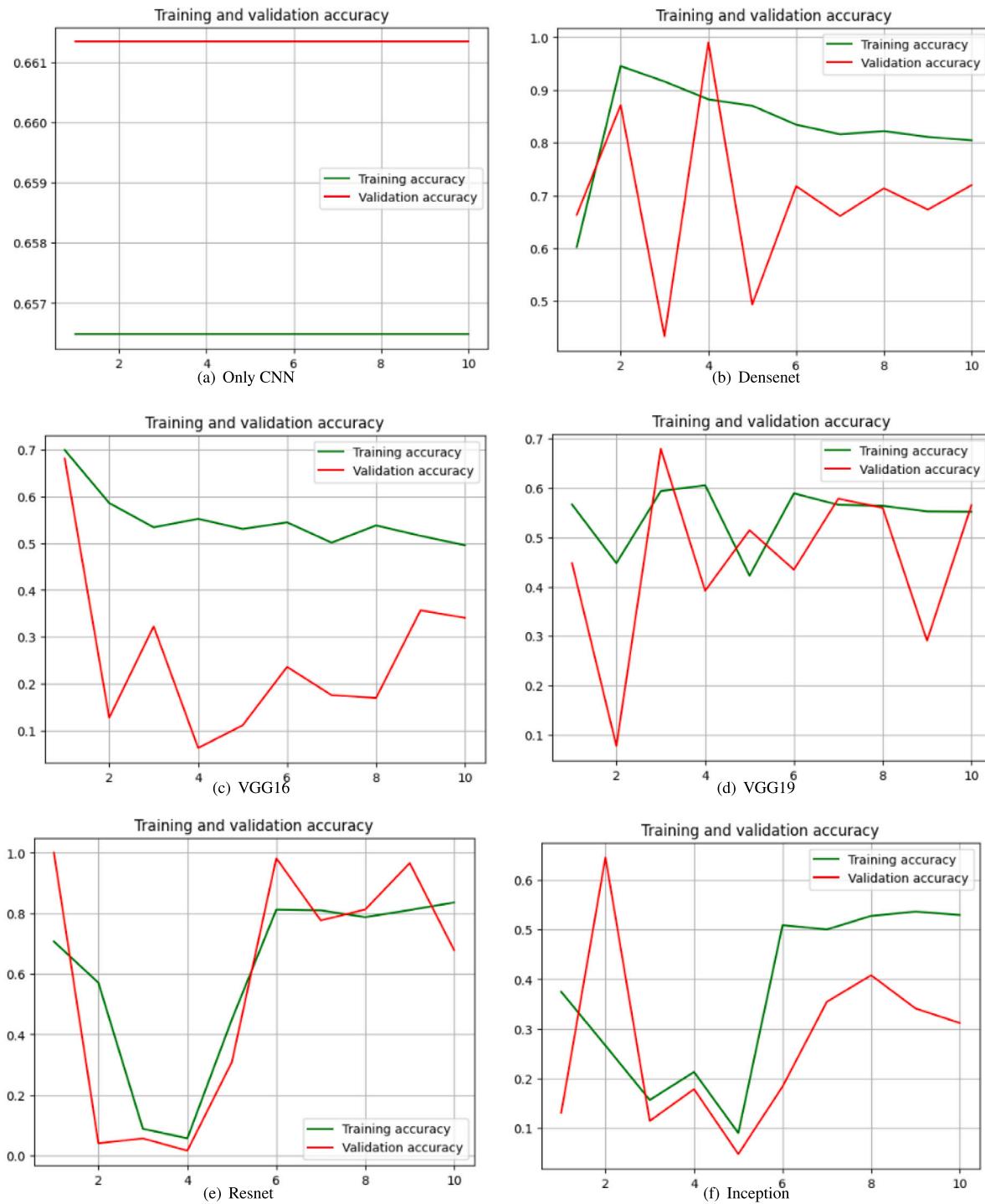


Fig. 16. Accuracy Analysis of Different Architectures for both Training and Testing.

architectures to analyze MRI images, with two approaches: replacing the last layer with an ANN and analyzing important weights using a LIME interpretable explainer. The second approach removes the last layer and uses pre-trained weights to predict new features, and lastly uses CNN architecture to detect Parkinson's disease. Our study also discovered that, for MRI image analysis, Densenet performed best in terms of accuracy reaching 73.66%, however, this requires the most time to execute. Furthermore, considering training time, VGG16 has minimal elapsed time at 98.481 s. Here, Table 13 below provides a summary of the models' running times.

While conducting this study, we faced a variety of challenges while performing our research. While conducting machine learning approaches we adopted 12 sets of datasets. Therefore, It was very difficult to handle because the dataset is highly imbalanced. Later on, while conducting Transfer learning approaches using MRI samples, we faced difficulties in creating an image dataset because numerous samples were found that were irrelevant to our studies. Therefore, we need to analyze every single image deeply before adding it to our dataset

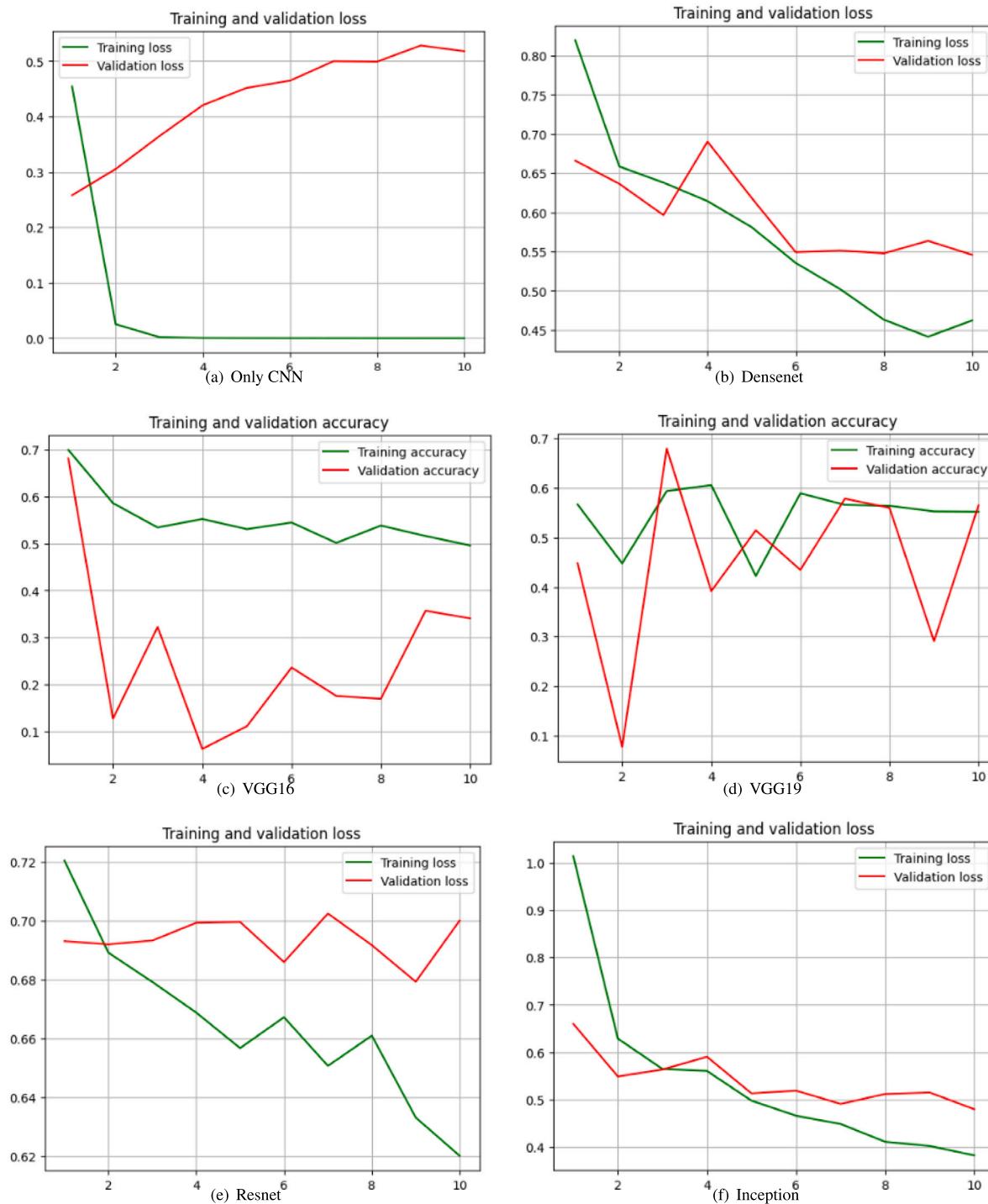


Fig. 17. Loss Analysis of Different Architectures for Both Training and Testing.

6. Broader impact

This comprehensive study extensively analyzes 12 distinct clinical assessment tests, aiming to pinpoint the key features associated with the development of Parkinson's disease. Consequently, it holds potential benefits for clinical experts and researchers. By offering insights into significant symptoms that patients may exhibit, it enables professionals to propose precise assessment tests for accurate diagnosis. Additionally, individuals diagnosed with Parkinson's can gain benefits from this

research, obtaining valuable insights into early signs and symptoms that can guide them toward appropriate treatment options.

7. Limitations

This study presents an in-depth analysis of disease by utilizing machine learning and transfer learning techniques on both clinical assessments and neuroimaging records. However, there are a few limitations associated with this work that we encountered.

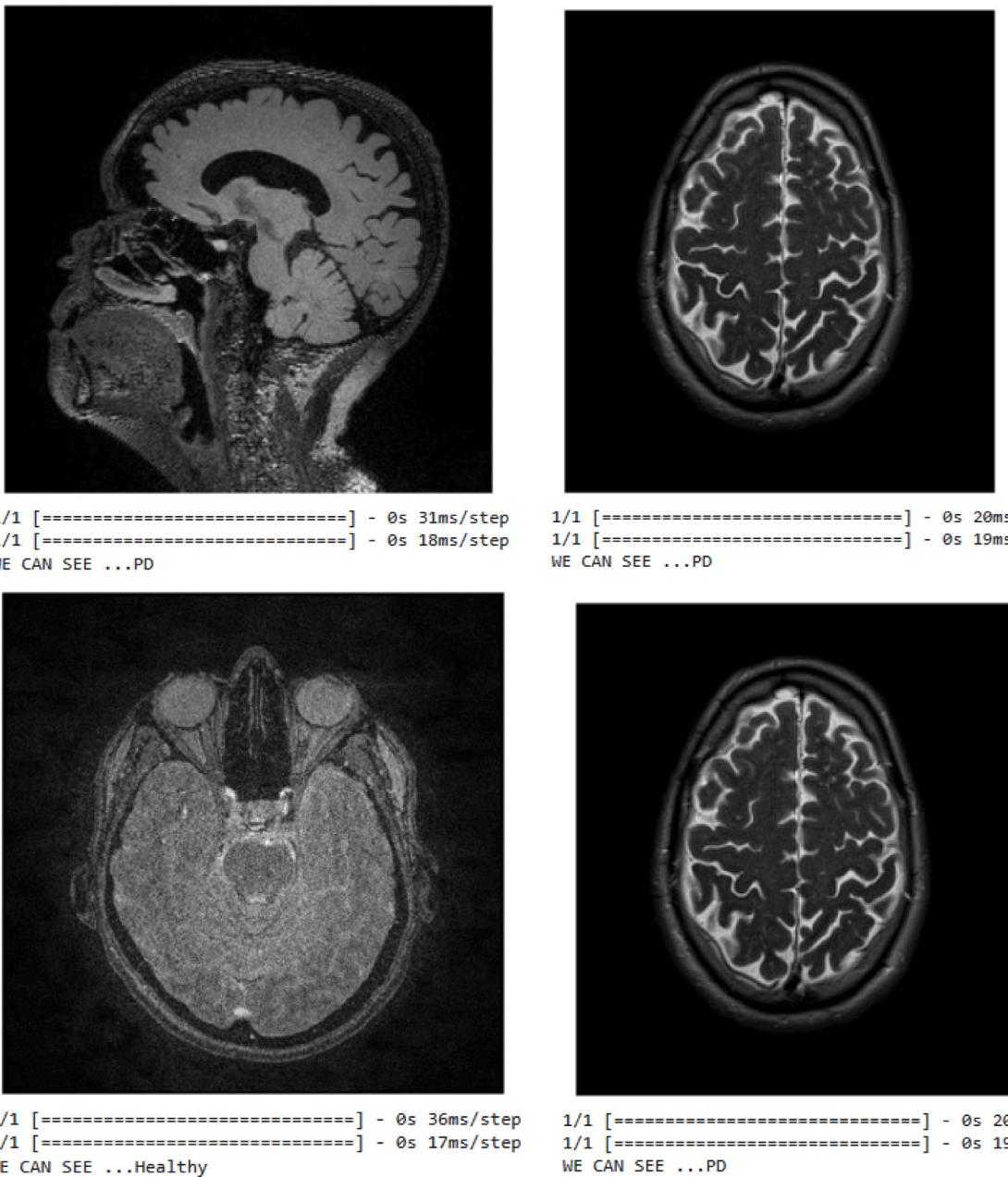


Fig. 18. Samples Predicted using The Best model.

- Our study faced limitations due to the unavailability of clinical experts, who could have contributed valuable insights and perspectives for improving our Parkinson's disease detection work.
- We encountered a significant volume of irrelevant data in our research, which restricted our capacity to build a larger and more representative dataset for our studies.

8. Conclusion

The objective of this study was to analyze Parkinson's disease by utilizing machine learning and transfer learning techniques on both clinical assessments and neuroimaging records. To achieve this goal, we collected both clinical assessment records and MRI samples from well-known PPMI database. To analyze the clinical assessment records, 4 different resampling methods are used and we identified the hybrid of SMOTE and ENN as the best method to handle imbalanced data. Additionally, three machine learning models were applied to train on these

resample records, and then precision and recall score was evaluated. Among the three algorithms, AdaBoost provided the best recall which is **89.58%** after using the SMOTE-ENN hybrid Technique. Therefore, it is determined that the hybrid of SMOTE and ENN is the most effective approach for addressing imbalanced data in our study. Afterward, 4 different feature selection methods are used to find out the top 10 features using these news samples. Finally, 9 different machine learning algorithms are used to train the samples of these 10 features and then validate the model. We also used explainable AI techniques like LIME and SHAP to interpret clinical assessment records. Hence, the Extra tree classifier outperforms the others in terms of accuracy reaching **98.44%** using the tree-based feature selection technique. In addition, to achieve higher accuracy, ET also demonstrated strong performance across other metrics, with precision, recall, and f1-score scores of 97.11%, 99.02%, and 98.06%, respectively. Apart from that, the study uses four pre-trained architectures to analyze MRI images, with two approaches. The first approach is to use the convolutional layer while replacing

Table 12

A comparison of related works and our work.

Ref	Data source	Sample size	Data type	Applied approach	Best result
Magesh et al. (2020)	PPMI	642	Image data	Applied transfer learning and VGG16 CNN architecture to accurately classify DaTscan SPECT images	Accuracy 95.20%
Prashanth et al. (2016)	PPMI	584	Tabular data	Applied Naïve Bayes, SVM, Boosted Trees, and RF.	SVM provides 96.40%
Arafe et al. (2023)	PPMI	144	Image data	Applied DT, Gaussian Naïve Bayes, KNN, SVM	AUC peak value 0.685%
Camacho et al. (2023)	PPMI, COMPASS-ND, BioCog, PD-MCI Calgary-C-Big, NEUROCON, Tao Wu, ON Japan, Hamburg, UK Biobank, OASIS3, SALD, PD-MCI Montreal	2041	Image data	Applied CNNJacobians, CNNIntensity, CNNCombined, and SVM-RBF Parkinson	CNN Jacobians accuracy 79.3%
Wang et al. (2023)	Ruijin Hospital, Hospital of Zhengzhou University	534	Image	Applied CA-Net, AG-SE-ResNeXt50, VGG16, ResNet50, SE-ResNeXt50	ROI Accuracy 98.70%
Veetil et al. (2021)	PPMI	242	Image	Applied VGG16, VGG19, ResNet50, Xception, and DenseNet201	VGG16, VGG19 accuracy of 92.60%
Praneeth et al. (2023)	PPMI	591	Image data	Apply AlexNet (19), DMVDA (20), DNN (21), SVM (18), Proposed DRCNN	DRCNN accuracy 98.87% (with optimization)
Tomer et al. (2022)	PPMI	968	DICOM data	Apply GLCM(angle = 0, 45, 90, 90), PCA	PCA accuracy 90.5%
Solayman et al. (2023)	Department of Health, Israel	2151 898	Numerical data	Apply SMOTE, Lime, KNN, RF, SVM, Decision Tree and LR	Accuracy 96.34%
Our work	PPMI	1277 & 2500	Clinical assessment data and MRI data	Using clinical assessment analysis, we Applied four resamplings, four feature selections and 9 different algorithms for training, and also Lime and Shap. Using MRI image analysis, apply two approaches using transfer learning models, custom ANN and CNN	Using clinical assessment analysis accuracy of 98.44% and using MRI samples accuracy of 85.08%

Table 13

The running time (in sec) of different transfer learning architectures.

Architectures	Features extraction time	Training time	Total required time
Densenet169	303.323	9.369	312.7
VGG16	89.12	9.361	98.481
VGG19	85.08	21.824	106.9
Resnet50	88.171	15.800	104.1
InceptionV3	106.004	13.978	119.982

the remaining two layers with our own customized ANN and analyze important weights using a LIME interpretable explainer. For Approach 1, the highest accuracy is achieved with InceptionV3, reaching 50.24%, accompanied by a loss of 0.76. The second approach removes the last layer and uses pre-trained weights to predict new features, and lastly, CNN architecture is utilized to detect Parkinson's disease. DenseNet169 achieved the highest accuracy of **85.08%**, which was accompanied by a loss of 0.44.

Looking ahead, we aspire to examine a larger and deeper dataset of MRI records and apply diverse deep-learning techniques to bolster the model's performance. Future research endeavors also seek to enhance the practical applicability and usability of the study through the development of a user-friendly interface or application designed specifically for healthcare professionals. Additionally, efforts will be made toward the implementation of real-time monitoring systems aimed at providing uninterrupted support and comprehensive care to patients.

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.

References

- Abdelkawy, M., Sabir, Z., Guirao, J., & Saeed, T. (2020). Numerical investigations of a new singular second-order nonlinear coupled functional Lane-Emden model. *Open Physics*, 18, 770–778, Crossref, ISI.
- Ahmad, S. O., Longhurst, J., Stiles, D., Downard, L., & Martin, S. (2023). A meta-analysis of exercise intervention and the effect on Parkinson's disease symptoms. *Neuroscience Letters*, 801, Article 137162.
- Almomani, A., Nahar, K., Alauthman, M., Al-Betar, M. A., Yaseen, Q., & Gupta, B. B. (2024). Image cyberbullying detection and recognition using transfer deep machine learning. *International Journal of Cognitive Computing in Engineering*, 5, 14–26.
- Arafe, M., Bhagwat, N. P., Chatelain, Y., Dugré, M., Sokolowski, A., Wang, M., et al. (2023). Predicting Parkinson's disease progression using MRI-based white matter radiomic biomarker and machine learning: A reproducibility and replicability study. *bioRxiv*, 2023-2005.
- Armstrong, M. J., & Okun, M. S. (2020a). Diagnosis and treatment of Parkinson disease: A review. *Jama*, 323(6), 548–560.
- Armstrong, M. J., & Okun, M. S. (2020b). Diagnosis and treatment of Parkinson disease: A review. *Jama*, 323(6), 548–560.
- Balnarsaiah, B., Nayak, B. A., Sujeetha, G. S., Babu, B. S., & Vallabhaneni, R. B. (2023). Parkinson's disease detection using modified ResNeXt deep learning model from brain MRI images. *Soft Computing*, 1–10.
- Bereczki, D. (2010). The description of all four cardinal signs of Parkinson's disease in a Hungarian medical text published in 1690. *Parkinsonism & Related Disorders*, 16(4), 290–293.
- Camacho, M., Wilms, M., Mouches, P., Almgren, H., Souza, R., Camicioli, R., et al. (2023). Explainable classification of Parkinson's disease using deep learning trained on a large multi-center database of T1-weighted MRI datasets. *NeuroImage: Clinical*, 38, Article 103405.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Chen, Q., Sabir, Z., Raja, M. A. Z., Gao, W., & Baskonus, H. M. (2023). A fractional study based on the economic and environmental mathematical model. *Alexandria Engineering Journal*, 65, 761–770.
- Connolly, B. S., & Lang, A. E. (2014). Pharmacological treatment of Parkinson disease: A review. *Jama*, 311(16), 1670–1683.
- Dąbrowska, M., Schinwelski, M., Sitek, E. J., Muraszko-Klaudel, A., Brockhuis, B., Jamrozik, Z., et al. (2015). The role of neuroimaging in the diagnosis of the atypical Parkinsonian syndromes in clinical practice. *Neurologia I Neurochirurgia Polska*, 49(6), 421–431.
- Erdaş, Ç. B., & Sümer, E. (2023). A fully automated approach involving neuroimaging and deep learning for Parkinson's disease detection and severity prediction. *PeerJ Computer Science*, 9, Article e1485.
- Evangelou, E., Tsianos, G., & Ioannidis, J. P. (2008). Doctors' versus patients' global assessments of treatment effectiveness: empirical survey of diverse treatments in clinical trials. *Bmj*, 336(7656), 1287–1290.

- Gazewood, J. D., Richards, D. R., & Clebak, K. (2013). Parkinson disease: An update. *American Family Physician*, 87(4), 267–273.
- Guerrero Sánchez, Y., Sabir, Z., Günerhan, H., & Baskonus, H. M. (2020). Analytical and approximate solutions of a novel nervous stomach mathematical model. *Discrete Dynamics in Nature and Society*, 2020.
- Gupta, P., Beylergil, S., Murray, J., Jacobs, J., Kilbane, C., Shaikh, A. G., et al. (2021). Effects of Parkinson disease on blur-driven and disparity-driven vergence eye movements. *Journal of Neuro-Ophthalmology*, 41(4), 442–451.
- He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)* (pp. 1322–1328). Ieee.
- İlhan, Ö., & Şahin, G. (2024). A numerical approach for an epidemic SIR model via Morgan-Voyce series. *International Journal of Mathematics and Computer in Engineering*.
- Islam, N., Turza, M. S. A., Fahim, S. I., & Rahman, R. M. (2024). Single and multi-modal analysis for Parkinson's disease to detect its underlying factors. *Human-Centric Intelligent Systems*, 1–19.
- Magesh, P. R., Myloth, R. D., & Tom, R. J. (2020). An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery. *Computers in Biology and Medicine*, 126, Article 104041.
- Nisar, K., Sabir, Z., Raja, M. A. Z., Ibrahim, A. A. A., Erdogan, F., Haque, M. R., et al. (2021). Design of Morlet wavelet neural network for solving a class of singular pantograph nonlinear differential models. *IEEE Access*, 9, 77845–77862.
- Parkinson, J. (2002). An essay on the shaking palsy. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 14(2), 223–236.
- Praneeth, P., Satyvika, M., Kommareddy, V., Sarath, M., Mallela, S., Vani, K. S., et al. (2023). Classification of Parkinson's disease in brain MRI images using Deep Residual Convolutional Neural Network (DRCNN). *Applied Computer Science*, 19(2).
- Prashanth, R., Roy, S. D., Mandal, P. K., & Ghosh, S. (2016). High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. *International Journal of Medical Informatics*, 90, 13–21.
- Rana, A., Dumka, A., Singh, R., Rashid, M., Ahmad, N., & Panda, M. K. (2022). An efficient machine learning approach for diagnosing Parkinson's disease by utilizing voice features. *Electronics*, 11(22), 3782.
- Sabir, Z., Raja, M. A. Z., Baskonus, H. M., & Ciaccio, A. (2023). Numerical performance using the neural networks to solve the nonlinear biological quarantined based COVID-19 model. *Atti Della Accademia Peloritana Dei Pericolanti-Classe Di Scienze Fisiche, Matematiche E Naturali*, 1(1), 10.
- Sabir, Z., Raja, M. A. Z., Guirao, J. L., & Shoib, M. (2021). A novel design of fractional Meyer wavelet neural networks with application to the nonlinear singular fractional Lane-Emden systems. *Alexandria Engineering Journal*, 60(2), 2641–2659.
- Sabir, Z., Raja, M. A. Z., Shoib, M., & Aguilar, J. G. (2020). FMNEICS: Fractional Meyer neuro-evolution-based intelligent computing solver for doubly singular multi-fractional order Lane-Emden system. *Computational & Applied Mathematics*, 39, 1–18.
- Sabir, Z., Wahab, H. A., Javeed, S., & Baskonus, H. M. (2021). An efficient stochastic numerical computing framework for the nonlinear higher order singular models. *Fractal and Fractional*, 5(4), 176.
- Sanchez, Y. G., Umar, M., Sabir, Z., Guirao, J. L., & Raja, M. A. Z. (2018). Solving a class of biological HIV infection model of latently infected cells using heuristic approach. *Discrete and Continuous Dynamical Systems S*, 14.
- Singh, R., Mishra, J., & Gupta, V. K. (2023). The dynamical analysis of a tumor growth model under the effect of fractal fractional Caputo-Fabrizio derivative. *International Journal of Mathematics and Computer in Engineering*.
- Solayman, S., Aumi, S. A., Mery, C. S., Mubassir, M., & Khan, R. (2023). Automatic COVID-19 prediction using explainable machine learning techniques. *International Journal of Cognitive Computing in Engineering*, 4, 36–46.
- Thakur, K., Kapoor, D. S., Singh, K. J., Sharma, A., & Malhotra, J. (2022). Diagnosis of Parkinson's disease using machine learning algorithms. In *Congress on Intelligent Systems* (pp. 205–217). Springer.
- Tomer, S., Khanna, K., Gambhir, S., & Gambhir, M. (2022). Comparison analysis of GLCM and PCA on Parkinson's disease using structural MRI. *International Journal of Information Retrieval Research (IJIRR)*, 12(1), 1–15.
- Umar, M., Sabir, Z., Raja, M. A. Z., Baskonus, H. M., Yao, S.-W., & İlhan, E. (2021). A novel study of Morlet neural networks to solve the nonlinear HIV infection system of latently infected cells. *Results in Physics*, 25, Article 104235.
- Usman, T. M., Saheed, Y. K., Ignace, D., & Nsang, A. (2023). Diabetic retinopathy detection using principal component analysis multi-label feature extraction and classification. *International Journal of Cognitive Computing in Engineering*, 4, 78–88.
- Veetil, I. K., Gopalakrishnan, E., Sowmya, V., & Soman, K. (2021). Parkinson's disease classification from Magnetic Resonance Images (MRI) using deep transfer learned convolutional neural networks. In *2021 IEEE 18th India council international conference* (pp. 1–6). IEEE.
- Wang, Y., He, N., Zhang, C., Zhang, Y., Wang, C., Huang, P., et al. (2023). An automatic interpretable deep learning pipeline for accurate Parkinson's disease diagnosis using quantitative susceptibility mapping and T1-weighted images. *Technical report*, Wiley Online Library.
- Yang, F., Wang, K., Sun, L., Zhai, M., Song, J., & Wang, H. (2022). A hybrid sampling algorithm combining synthetic minority over-sampling technique and edited nearest neighbor for missed abortion diagnosis. *BMC Medical Informatics and Decision Making*, 22(1), 344.
- Yen, C., Lin, C.-L., & Chiang, M.-C. (2023). Exploring the frontiers of neuroimaging: A review of recent advances in understanding brain functioning and disorders. *Life*, 13(7), 1472.