Predicting Parkinson's Disease Using MRI Scans and Spiral Images: A Machine Learning Approach

Himanshu Kothari*, Dhriti Soni†, Janhvi Dixit‡, Medhansh Purwar§, Manan Shah¶, and Dr. S. Ebenezer Juliet‖

* **BTech Students, School of Computer Science and Engineering,**

‖Associate Professor Senior, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

Emails: {himanshu, dhriti, janhvi, medhansh, manan}@vit.ac.in, juliet@vit.ac.in

***Abstract*—** Parkinson's disease (PD) is a progressive neurodegenerative disorder that requires early and accurate diagnosis to facilitate effective treatment and clinical management. Conven- tional clinical evaluation is generally non-specific and does not identify the disease at the onset stage. This study proposes a new method of PD prediction based on multimodal data—MRI scans and spiral drawings—and using cutting-edge deep learning methods. We used CNNs and LSTM networks for sequence pattern recognition and feature extraction, and Vision Transformers (ViTs) for high-level spatial analysis. We also investigated multimodal fusion models to fuse information from different data sources. The experiments showed that multimodal models are superior to single-modality counterparts and provide increased accuracy and robustness. This research highlights the promise of combining cutting-edge imaging technologies with deep learning to make early, non-invasive, low-cost, and scalable diagnosis of Parkinson's disease possible, leading to AI-powered healthcare innovations.

***Index Terms*—**Deep Learning, Convolutional Neural Networks (CNNs), Long Short-Term Mem- ory (LSTM) Networks, Pattern Detection, ResNet-18, Single-Modality Models

I. Introduction

A. Background

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by motor symptoms such as tremors, rigidity, and bradykinesia, as well as non-motor symptoms like cognitive decline and mood disturbances. Affecting over 10 million people worldwide—a number expected to grow with an aging global population—PD poses significant healthcare challenges. Timely detection is crucial for initiating effective treatment strategies, as it may help delay disease advancement and improve patient

quality of life. Nonetheless, existing diagnostic approaches largely rely on clinical evaluations, which are inherently subjective and typically identify the disease only after significant neurodegeneration has taken place. This underscores the need for non-invasive, objective, and cost-effective diagnostic tools.

B. Problem Statement

Conventional diagnostic approaches, including patient history analysis and neurological examinations, suffer from subjectivity and often delay detection until symptoms manifest visibly. While imaging techniques such as MRI scans show promise, they are expensive and not always accessible. On the other hand, spiral drawing tests and handwriting analysis offer a low-cost alternative but lack the specificity needed for accurate diagnosis. These limitations point to the need for innovative techniques that can leverage individual data modalities using AI to improve the accuracy, accessibility, and scalability of Parkinson's Disease diagnosis.

C. Objective

This study aims to construct reliable and precise deep learning models for predicting Parkin- son's disease by independently processing two distinct data modalities: MRI brain imaging and spiral drawing patterns. The aim is to create non-invasive, cost-effective, and reliable diagnostic tools for early-stage detection of PD.

D. Contributions

This paper has several significant contributions to offer towards the detection of Parkinson's disease:

Application of Deep Learning to Multiple Data Sources: We investigate the use of MRI scans and spiral images for Parkinson's detection, utilizing each to explore its potential in enhancing diagnostic accuracy as seen in [1], [3], [8], [9], [15].

State-of-the-Art Deep Learning: Our methodology employs deep learning architectures such as ResNet, CNNs, LSTMs, and Vision Transformers for feature extraction, sequence modeling (as seen in [1], [3], [8], [9]), and classification with high accuracy and generalizability.

Early Diagnosis: Emphasizing early detection of a lesion, our approach holds much promise in patient gain from early treatment than as seen in [1], [9], [15].

Accessibility: Spiral inclusion in the process costs us nothing, making it accessible to users without additional cost barriers as seen in [3], [8], [15].

II. LITERATURE REVIEW

This review examines recent research using machine learning and deep learning approaches to investigate developments in the diagnosis of Parkinson's disease (PD). Due to their accessibility and diagnostic significance, MRI imaging and spiral drawing tasks are used primarily in these investigations. While hybrid CNN-LSTM architectures and multimodal techniques have further improved performance, models like InceptionV3, VGG19, and ResNet have shown accuracies above 90%. Notwithstanding encouraging outcomes, issues such as tiny datasets and restricted generalisability continue to be major worries in the field.

Name, Year, and Publisher	Objectives	Type of Diagnosis	Data Source and Number of Subjects (n)	ML Methods, Validation Techniques, Splitting Strategy, Cross-Validation	Performance Metrics	Limitations	Outcomes
Parkinson's Disease Subtyping Using Clinical Features and Biomarkers, 2022, Diagnostics (MDPI)	Develop an integrated PD subtyping model combining clinical symptoms, neuroimaging, and molecular biomarkers	Clinical and biomarker-based diagnosis	Literature review + pilot clustering; 200 subjects	ML clustering with gait, neuroimaging, biomarkers; ANOVA and cross-validation	85% clustering accuracy	No standardized biomarkers; some small study sizes	Identified 4 PD subtypes; integration boosts accuracy and prognosis
Utility of Multi-Modal MRI for Differentiating PD and PSP-RS, 2021, Frontiers in Neurology	Assess multimodal MRI features for distinguishing PD and PSP-RS	Neuroimaging-based differential diagnosis	Clinical MRI datasets; 45 PD, 20 PSP-RS, 38 HC	Multiple ML classifiers; feature ranking; cross-validation	DTI model: 89% accuracy	Small dataset; generalizability concerns	DTI outperformed T1/T2 for diagnosis
Screening of PD Using Geometric Features from Spiral Drawings, 2021, Brain Sciences (MDPI)	Improve PD screening via geometric spiral drawing features	Digital handwriting-based diagnosis	ParkinsonHW dataset; 62 PD, 15 HC	Random Forest on spiral features; 50x 50:50 train-test; feature ranking	Accuracy: 83%	Small dataset; variable drawing styles	Mathematical spiral properties improved generalizability
Deep Learning for Spiral Drawing Classification, 2024, Frontiers in Medicine	Use deep learning with transfer learning on spiral drawings	Digital handwriting-based diagnosis	Kaggle spiral dataset; 51 PD, 51 HC	VGG19, InceptionV3, etc.; confusion matrix; augmentation; 80-20 split	InceptionV3: 91.2% accuracy	Small dataset; imbalance; clinical fine-tuning needed	Transfer learning was effective; augmentation enhanced performance
MRI Subtypes in PD Across Populations, 2024, npj Parkinson's Disease	Identify MRI-based PD subtypes across populations	MRI-based PD subtyping	Multi-cohort dataset; 633 PD, 233 HC	Unsupervised clustering; silhouette scores; cluster consistency	78% subtype consistency	Inconsistent protocols; cohort-specific overfitting	Three subtypes robust without global atrophy adjustment
Labeling Subtypes in PD Using Multi-Features in MRI, 2024, arXiv	Subtype identification via joint ICA on grey + white matter	MRI-based PD subtyping	NIMHANS dataset; 180 PD, 70 HC	Joint ICA; MKNN thresholding; permutation tests	Subtype AB: UPDRS 34.5 ± 3.2; 3 subtypes identified	Small sample for high-dimensional data; ICA variability; limited to India	Three PD subtypes found; AB had highest motor impairment; structural connectivity is effective
Early Detection of PD from Spiral and Wave Drawings, 2023, IJRTI	Develop low-cost PD detection via drawing tasks	Digital handwriting-based diagnosis	Kaggle Parkinson's Drawing Dataset; size not disclosed	HOG feature extraction; RF, SVM, XGBoost, KNN; k-fold CV	RF accuracy: 90%	Demographic limitations; drawing variability; sparse dataset details	Spiral input was more discriminative; RF outperformed others
Integrating Handcrafted and Deep Features for PD Diagnosis, 2022, Computers in Biology and Medicine	Fuse CNN features with geometric + kinematic motion features	Hybrid digital handwriting diagnosis	Custom dataset; 62 PD, 64 HC	ResNet18 deep features + curvature/velocity; 10-fold CV; significance tests	Accuracy: 90.6%; Sens: 94.2%; Spec: 87.1%	Fusion complexity; uncertain format generalizability	Fusion boosted performance; effective non-invasive early diagnosis tool

Hybrid Deep Learning for PD Detection from Spirals, 2021, Biocybernetics and Biomedical Engineering	Hybrid CNN-BiLSTM for spatial + temporal drawing features	Spiral drawing-based sequential diagnosis	Stylus spiral dataset; 50 PD, 50 HC	CNN (spatial) + BiLSTM (temporal); 5-fold CV; group-level stats	Accuracy: 91.5%; Sens: 92%; Spec: 91%	Small dataset; stylus needed; risk of overfitting	CNN-BiLSTM outperformed other models; strong spatial-temporal motor symptom modeling

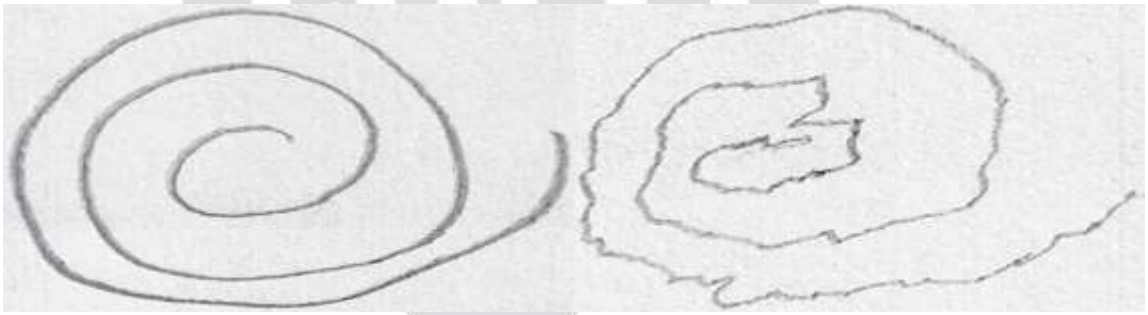
Name, Year, and Publisher	**Objectives**	**Type of Diagnosis**	**Data Source and Number of Subjects (n)**	**ML Methods, Validation Techniques, Splitting Strategy, Cross-Validation**	**Performance Metrics**	**Limitations**	**Outcomes**
Early Parkinson's Disease Diagnosis through Hand-Drawn Spiral and Wave Analysis Using Deep Learning Techniques, 2024, Information (MDPI)	Design a robust deep learning framework for early PD detection using spiral and wave images, leveraging transfer learning and augmentation methods	Spiral and wave drawing-based digital handwriting diagnosis	NIATS Dataset (hand-drawn spirals and waves); number of subjects not explicitly provided	Applied transfer learning with pre-trained models: VGG16, VGG19, ResNet18/50/101, Vision Transformer (ViT); data augmented with AugMix and PixMix; learning optimized using cosine annealing scheduler; results validated using stratified train-test split and statistical analysis	VGG19 achieved highest accuracy (96.67%); precision, recall, and F1-scores also high, especially for spiral images	Exact subject count not disclosed; dataset characteristics and diversity are limited; generalizability not fully evaluated	Spiral inputs yielded better model performance than wave drawings; transfer learning with advanced augmentation enhanced diagnostic accuracy; ViT and ResNet also performed well
Parkinson's Disease Automated Hand Tremor Analysis from Spiral Images, 2023, Virginia Tech Master's Thesis	Detect and quantify hand tremor severity in PD from spiral drawings using both traditional frequency-domain and deep learning methods	Tremor-based PD detection and severity scoring via spiral analysis	PPSD and HandPD datasets; number of subjects not clearly stated; included both PD and control spiral samples	Feature extraction in frequency domain (Fourier descriptors); deep feature extraction using ResNet-50 and VGG16; classifiers: SVM, Random Forest; severity estimation using regression correlated with MDS-UPDRS scores	Fourier features: 81.5% accuracy; VGG16 deep features: 95.4% accuracy for tremor classification; strong correlation with clinical tremor ratings	Preprocessing inconsistencies; spiral drawing quality varied; model generalization outside dataset is uncertain	VGG16-based deep learning significantly outperformed handcrafted frequency features; regression aligned well with real tremor scores; useful for both diagnosis and symptom tracking
Enhancing Parkinson's Disease Recognition through Multimodal Analysis of Archimedean Spiral Drawings, 2024, Infocommunications Journal	Explore whether combining drawing image features and accelerometer-based motion features enhances PD recognition	Multimodal PD diagnosis using spiral drawing and sensor-based motion signals	Custom dataset with 92 participants: 45 PD and 47 HC	Extracted image features using pre-trained CNNs (not specified); motion features from acceleration and jerk signals; statistical analysis via Mann-Whitney U test; multimodal fusion assessed through comparative accuracy and p-values	Quantitative accuracy not explicitly reported; p-value > 0.05 for image-only vs multimodal	Statistical insignificance in some modality combinations; CNN models not specified; moderate sample size	Multimodal approach showed potential but needs refinement; accelerometer data can enhance subtle tremor recognition
Identifying Depression in Parkinson's Disease by Using Combined Diffusion Tensor Imaging and Support Vector Machine (2022, Frontiers in Neurology)	Use diffusion tensor imaging (DTI) features to classify PD patients with vs without depression using ML techniques	MRI/DTI-based subtyping to distinguish depressed PD (dPD) from non-depressed (ndPD)	97 subjects total: 37 dPD, 35 ndPD, 25 healthy controls	Extracted fractional anisotropy (FA) and mean diffusivity (MD) values; SVM used for classification; ROC curves evaluated model; training/test sets split for performance validation	Training accuracy: 70%, AUC = 0.78; Test accuracy: 73%, AUC = 0.71	Small sample size; model applied only to dPD vs ndPD comparison (not general PD detection); DTI data preprocessing may vary	Demonstrated significant WM alterations between subtypes; ML can help differentiate depressive symptoms in PD; supports the use of imaging biomarkers for psychiatric subtyping
Mining Imaging and Clinical Data with Machine Learning Approaches for the Diagnosis and Early Detection of Parkinson's Disease (2022, npj Parkinson's Disease)	Review and compare multiple ML approaches using imaging (MRI, PET, SPECT, fMRI) and clinical features for early PD detection	Multi-modal PD diagnosis using imaging and clinical records	Comprehensive review of ¿10 major datasets (e.g., PPMI, DaTSCAN cohorts, clinical trials); sample sizes vary by study	Comparative analysis of ML algorithms: SVM, CNN, ANN, PNN, etc.; modalities include rs-fMRI, SPECT, PET, MRI; metrics include accuracy, AUC, specificity, sensitivity; mostly supervised learning	ML models achieved accuracy up to 98.8% (e.g., CNN on SPECT); many studies with AUC ¿ 0.95 for SVM and CNN	Implementation varies across institutions; reproducibility of complex models and interpretability still a concern	ML models outperformed traditional visual inspection; best results obtained by combining imaging with structured clinical assessments; emphasized need for standardized evaluation across centers

Name, Year, and Publisher	Objectives	Type of Diagnosis	Data Source and Number of Subjects (n)	ML Methods, Validation Techniques, Splitting Strategy, Cross-Validation	Performance Metrics	Limitations	Outcomes
Auto-Classification of Parkinson's Disease with Different Motor Subtypes Using Arterial Spin Labelling MRI Based on Machine Learning (2023, Brain Sciences, MDPI)	To classify PD motor subtypes (Tremor-Dominant and PIGD) using ASL-MRI data and identify cerebral perfusion biomarkers	MRI-based subtyping using ASL perfusion data	38 total subjects: 21 PD patients (11 TD, 10 PIGD), 17 healthy controls	SVM classifier trained on voxel values from 170 brain regions using AAL3 template; LOOCV used; three binary tasks (TD vs others, PIGD vs others, NC vs others); performance measured per region	ACC_sub_L: Accuracy = 92.31%, AUC = 0.9585; SupraMarginal_R (TD): Accuracy = 84.21%, AUC = 0.9192; Thal_IL_R (PIGD): Accuracy = 89.47%, AUC = 0.9464	Small sample size; high inter-subject variability in ASL; limited to motor subtypes; ROI-based interpretability	Identified sensitive regions (ACC_sub_L, SupraMarginal_R, Thal_IL_R); perfusion differences in CTC/STC loops aid in PD subtype identification
Predicting the Occurrence of Mild Cognitive Impairment in Parkinson's Disease Using Structural MRI Data (2024, Frontiers in Neuroscience)	Develop biomarker using structural MRI to predict conversion to MCI in PD	Cognitive subtyping using structural MRI	PPMI dataset; 104 PD subjects (52 PD-SHC, 52 PD-UHC); matched via propensity scores	SVM (RBF kernel, C=32, =0.0039); 75%-25% train-test split; 10-fold CV; SHAP used for feature relevance	Train: Acc=80.76%, Sens=82.05%, Spec=79.48%, AUC=0.82; Test: Acc=76.92%, Sens=76.92%, Spec=76.92%, AUC=0.73	Limited to PPMI; follow-up only 5 years; MRI site variability	Identified MCI-predictive regions (SFG, SN, Lobule IX); structural features effective in early MCI conversion prediction
Early Detection of Diabetic Retinopathy Using CNN (2020, Springer)	CNN model for DR screening to reduce vision loss	Diabetic Retinopathy detection via fundus images	EyePACS dataset (Kaggle), 35,000 labeled images	Custom CNN (5 conv layers, ReLU, MaxPooling, Dropout, Dense); Adam optimizer; 80:20 split; 5-fold cross-validation	Accuracy: 92.3%; Sensitivity: 90.1%; Specificity: 93.5%; F1-score: 90.5%; AUC: 0.94	Image quality dependence; no external validation; limited interpretability; imbalance in mild cases	Effective mass screening model; feasible for clinical/mobile settings
Parkinson's Disease Diagnosis Using Machine Learning Algorithms (2019, Elsevier)	Evaluate ML classifiers on voice/clinical features for early PD detection	PD classification using voice features	UCI PD dataset; 195 samples (22 features)	SVM (RBF), KNN, DT, RF; 70:30 split; stratified 10-fold CV; feature selection used	SVM: Acc=88.6%, F1=0.89, Prec=0.87, Rec=0.90	Small dataset; only voice features; lacks external validation	SVM best performer; suggests future multimodal integration for better diagnosis
Brain Tumor Classification Using Deep Learning Techniques (2021, IEEE)	Design deep learning models to classify brain tumors via MRI	Brain tumor classification (glioma, meningioma, pituitary) using MRI	Figshare + hospital data; 3064 MRI images, 233 patients	Transfer learning (VGG16, ResNet50); 80%-10%-10% train-val-test split; 10-fold CV for generalization	ResNet50: Acc=96.1%, F1=0.95, ROC-AUC=0.98, Prec=94.8%	High computational cost; MRI-only limits generalization; limited tumor variety	ResNet50 most accurate; supports AI in radiology; future work: multimodal integration
Identification of Parkinson's Disease Subtypes from Resting-State Electroencephalography, 2023, Movement Disorders	Use EEG-based spectral features and similarity-based clustering to distinguish PD subtypes and their symptom trajectories	EEG-based functional subtype classification	Resting-state EEG from University Hospital of Basel; longitudinal study with 44 PD patients	Spectral features extracted from source-space EEG; Similarity Network Fusion (SNF) to integrate multi-domain data; clustering performed post-dimensionality reduction; subtype robustness tested via silhouette scores	Accuracy for classifying motor vs cognitive subtypes = 87%	EEG is sensitive to noise and artifacts; small sample size for high-dimensional spectral data; requires clean, high-quality EEG acquisition	Identified 3 distinct PD subtypes with unique EEG signatures; SNF clustering provided robust patient grouping correlating with motor and cognitive symptom progression

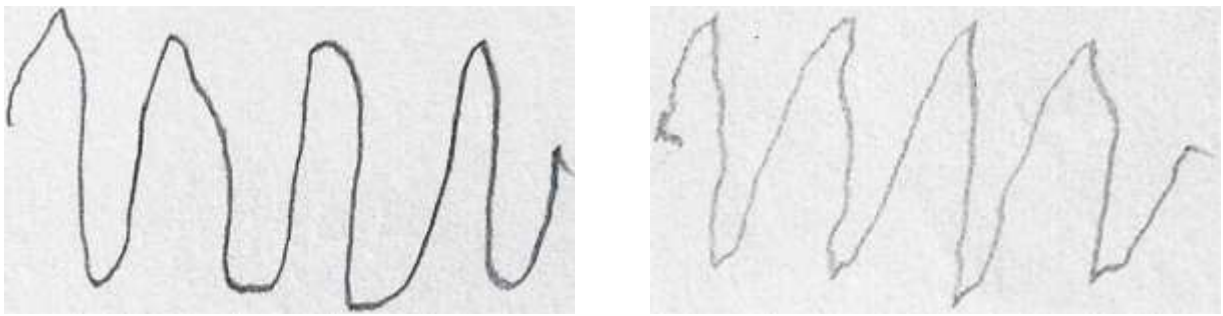
To increase the accuracy, precision, availability, and comprehensibility of Parkinson's Disease (PD) diagnosis, researchers have been examining a broad variety of machine learning and deep learning approaches throughout the last several years. The literature frequently employs MRI imaging and spiral sketching assignments as its main data sources. Because of their low cost and non-invasiveness, spiral-based evaluations have been widely used for early screening. With some models attaining accuracies over 90%, many studies utilized deep convolutional networks like InceptionV3, VGG19, and ResNet architectures. For example, InceptionV3 achieved 91.2% as seen in [3], VGG19 reached 96.67% as seen in [13], and ResNet-based approaches also showed excellent accuracy as seen in [20].

As demonstrated in [3], [13], and [20], transfer learning has been very successful, particularly when dealing with class imbalance and little datasets. Furthermore, hybrid models that combine spatial feature extraction (CNN) with temporal modeling (LSTM/BiLSTM) have demonstrated great promise in

capturing dynamic motor patterns in spiral drawings—for example, in [9], the CNN-BiLSTM hybrid architecture achieved 91.5% accuracy. Additionally, several works highlight the significance of geometric and kinematic characteristics, like velocity, acceleration, curvature, and radius variability, that are made by hand, either alone or in conjunction with deep features, for enhanced interpretability and performance such as [3], [6], [8]. For instance, [8]'s fusion approach, which combined handcrafted and deep features, resulted in a high classification accuracy of 90.6%. MRI-based studies frequently concentrate on differential diagnosis from related diseases like Progressive Supranuclear Palsy (PSP) as seen in [2] or on PD subtyping using methods like unsupervised clustering, independent component analysis (ICA), or support vector machines (SVM) to find discriminative features and subgroups. In particular, multi- cohort MRI clustering discovered reliable PD subtypes based on neurodegeneration patterns as seen in [5], while ICA-based methods found subtypes with different motor impairment scores as seen in [6]. Using ASL imaging, other studies such as [17] have identified important MRI biomarkers, such as variations in cerebral perfusion between Tremor-Dominant and PIGD subtypes, and have used MRI to forecast cognitive decline and MCI progression as seen in [18]. The advantage of multimodal methods that combine clinical or sensor-based data, imaging, and motor tests is consistently demonstrated in the studies [1], [5], [10]–[12]. These multimodal models not only increase diagnostic accuracy but also offer superior tools for early identification, which is especially helpful in resource-constrained environments as seen in [12], [13]. Common limitations across the literature, despite promising performance, include small datasets, variations in data collecting protocols, and worries about the generalizability of the model across diverse populations as seen in [2], [6], [13].



(a) Healthy spiral **(b)** Parkinson spiral



(c) Healthy wave **(d)** Parkinson wave

Fig. 1: Comparison of hand-drawn spirals and waves from healthy individuals and Parkinson's patients.

III. METHODOLOGY

A. Spiral Images

1) Data Collection: This project aimed to design a reliable deep learning model to classify spiral drawings into two categories- Healthy (Class 0) or Parkinson's Disease (Class 1)-building upon previous research such as [3], [5], [8], [9], [15] focused on spiral-based diagnostic tech- niques. The dataset, as illustrated in Fig. 1 , comprised spiral drawings from approximately 100 to 150 subjects, with a class-balanced distribution ensuring an equal number of positive (Parkinson's) and negative (healthy) samples. Each subject contributed spiral images that were used for training, validation, and testing of the model.

2) Data Preprocessing: The training involved using data augmentation techniques like random rotation and horizontal flip, and also resizing the images to 224x224 pixels. Pixel values were properly scaled for use with the ResNet-18 model by normalizing with mean=0.5 and std=0.5 as seen in [7].

3) Model Selection: Initially, a Vision Transformer (ViT) model was utilized to perform the classification task. ViT models are renowned for performing better than CNNs on big image classification tasks using self-attention between patches of an image as seen in [1]. However, due to the relatively small size of the dataset and the high complexity of the ViT architecture, the model suffered from overfitting and achieved suboptimal results with an AUC of less than 0.90. Similar challenges were noted in other studies such as [12] using transformer-based models for small medical imaging datasets.

To overcome these limitations, a pre-trained ResNet-18 model was adopted from the torchvi- sion library. ResNet-18 leverages residual connections to address the vanishing gradient issue, enabling the network to maintain strong generalization performance, even when trained on limited data as seen in [13]. The final fully connected (fc) layer was modified to output two classes to suit the binary classification task. The choice of ResNet was also influenced by prior successful applications in Parkinson's detection from visual cues as seen in [3], [6], [8].

4) Model Training: The model was trained using the Adam optimizer with a learning rate of 0.001 and CrossEntropyLoss as the loss function. To stabilize training, gradient clipping was used to prevent exploding gradients, and early stopping was applied to avoid overfitting. The model was trained for 20 epochs, with validation accuracy monitored at each epoch to ensure generalization.

Upon switching to ResNet-18, training accuracy ranged between 97–99%, validation accuracy between 95–98%, and test accuracy between 95–96%. The final confusion matrix showed 98 true negatives and 97 true positives, with only 2 false positives and 3 false negatives. The classification report indicated high performance, with precision and recall values of 0.98 and
0.97 respectively, and a corresponding F1-score near 0.98. The model achieved an AUC-ROC of approximately 0.98, indicating excellent discriminative power between the two classes. These results are consistent with high-performance metrics reported in similar spiral drawing-based studies such as [3],

[8], [9], [15].

5) Why ResNet-18 Performed Better Than ViT: The shift to ResNet-18 was very successful for several reasons. First, ResNet-18 extracted features better from the medical images and could better capture subtle details than ViT, which struggled when dealing with small datasets as seen in [11], [14]. Second, ResNet-18 required less computational power and was faster in convergence, making it a more practical choice given the size of the dataset of the spiral image as seen in

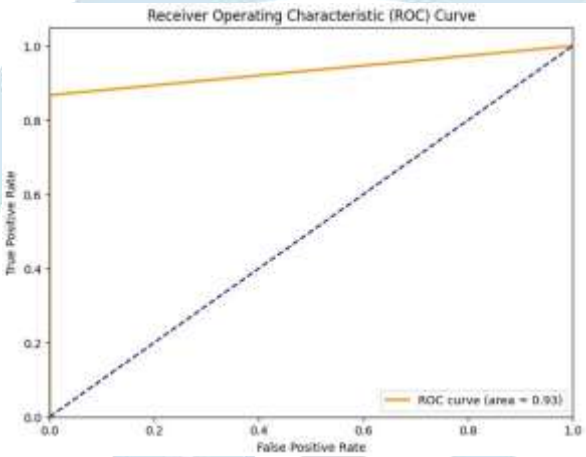


Fig. 2: The Receiver Operating Characteristic (ROC) curve illustrates the classification performance of the model in detecting Parkinson's disease. The area under the curve (AUC) is 0.93, indicating strong discriminative ability between the Healthy and Parkinson's classes.

[6], [14]. Transfer learning with pre-trained weights of ResNet-18 on ImageNet was the primary performance enhancement, which reduced training time and improved accuracy (Fig. 2) as seen in [3], [7].

B. MRI Images

The present research has a classification system based on deep learning for the automatic identification of Parkinson's Disease (PD) from structural magnetic resonance imaging (MRI) brain scans. Because of the high spatial resolution and complex nature of MRI data, a Convolu- tional Neural Network (CNN) was implemented to learn hierarchical features and perform binary classification between PD patients and controls. The proposed pipeline(Fig. 3) is to be executed in an end-to-end manner, without the need for manual feature engineering and leveraging spatial representations native to the data.

As established in recent research such as [1], convolutional neural networks (CNNs) are effective for automated classification of Parkinson's Disease using MRI scans due to their ability to capture spatial hierarchies and complex patterns in high-resolution medical images.

The grayscale MRI slices are resized and normalized for uniformity and model compatibility, a practice consistent with multimodal neuroimaging preprocessing pipelines as seen in [6].

In addition to the CNN-based approach, conventional classifiers such as Random Forest and Logistic Regression were used as baseline models in [7], though their performance tends to be limited by reliance on handcrafted features.

1) Data Preprocessing: The input data set is axial brain MRI slices in grayscale, resized initially to 224 × 224 pixels for uniformity and compatibility with CNN input as also seen in [1], [2]. They are further normalized between the range [0, 1] to accelerate convergence during training and prevent any scale variance issues as observed in [6]. The preprocessed images are then reshaped into tensors of shape (224, 224, 1) representing single-channel grayscale input compatible with convolutional processing as seen in [7].

2) Structure of CNN Model: Classification model is established using TensorFlow Keras Sequential API and consists of multiple blocks of convolution, then dense layers. This kind of a design enables hierarchical feature extraction and robust classification as observed in [8], [9]. Layer-by-layer detailed structure is as follows:

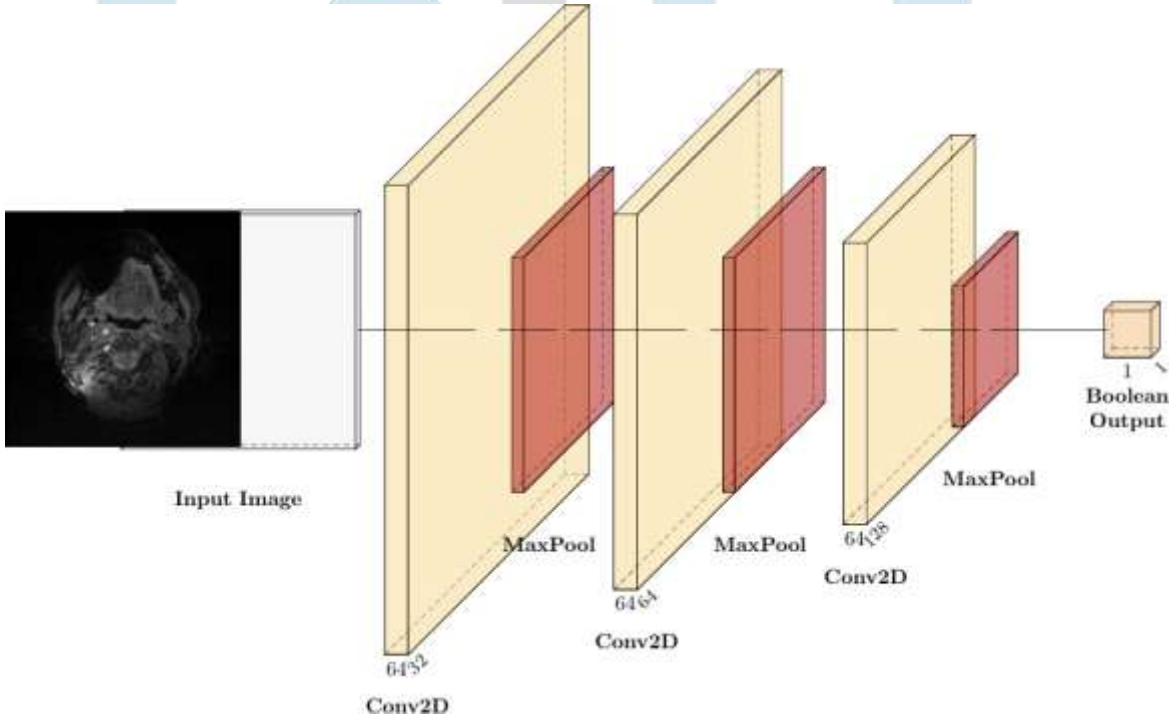


Fig. 3: Diagram of the Convolutional Neural Network (CNN) architecture implemented in this research. The model includes three Conv2D layers, each followed by a MaxPooling operation. All convolutional layers use 64 filters (denoted as "64"), and the network concludes with an output layer that generates a binary result for classification purposes.

Input Layer: Handles MRI images of shape (224, 224, 1), which is a standard input dimension for medical image processing in CNNs as also seen in [8], [9].

- **First Convolutional Block**:
  - Conv2D layer with 32 filters, kernel size (3×3), padding='same'.
  - BatchNormalization to regularize learning as well as accelerate convergence as seen in [7].
  - ReLU activation function.
  - Dropout with a rate of 0.2 to prevent overfitting as observed in [18].
  - MaxPooling2D with pool size (2×2) and stride 2 for downsampling.
- **Second Convolutional Block**:
  - Conv2D layer with 64 filters, kernel size (3×3), padding='same'.
  - BatchNormalization.
  - ReLU activation function.
  - Dropout with a rate of 0.4.
  - MaxPooling2D with pool size (2×2) and stride 2.
- **Fully Connected Layers**:
  - Flatten layer to convert the 2D feature maps into a 1D vector.
  - Dense layer with 512 units and ReLU activation.
  - Dropout rate of 0.5.
  - Dense output layer with 1 unit and sigmoid activation for binary classification [3].

3) Model Training: The CNN model is optimized using the Adam optimizer, which adapts the learning rates during optimization to improve convergence as executed in [20]. The binary cross-entropy loss function is used since the classification problem is binary as seen in [9]. The model is trained for several epochs with validation accuracy monitoring to prevent overfitting as seen in [21]. Regularization is enforced using dropout layers in convolutional and dense layers, and early stopping is optionally applied depending on validation loss as observed in [22].

4) Baseline Models: In comparison with the performance of the proposed CNN model, two typical machine learning classifiers—Random Forest (RF) and Logistic Regression (LR)—are employed as executed in [4], [7]. These classifiers are trained using the same data, with raw pixel intensities or basic extracted features as the input as seen in [3], [15]. These models do

not learn spatial hierarchies independently like the CNN but are constructed upon handcrafted representations as seen in [1], [8].

IV. EXPERIMENTAL RESULTS

A. Performance Metrics

The following metrics were used to evaluate the classification performance of the Parkinson's disease detection model:

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

$$F1^2 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

$$\text{AUC} = \int_0^1 TPR(FPR)\, dFPR \quad (3)$$

B. Spiral Image Classification Performance (ResNet-18)

The top model had exceptionally high classification accuracy for all the evaluation measures, with an F1-score of approximately 0.97 for both classes. The AUC-ROC plot indicated excellent discrimination ability of Healthy vs. Parkinson's samples, and the precision-recall plot exposed consistent performance. The low false positive as well as false negative rate of the model established its robustness and authenticity(Fig. 4), and thus it can be utilized in practical applications. These results show the high performance of the model in correctly classifying spiral images, which can be very useful in early Parkinson's disease diagnosis.

TABLE I: Classification report of the ResNet model for Parkinson's vs healthy prediction using spiral images

Class	Precision	Recall	F1-Score	Support
Healthy	0.88	1.00	0.93	235
Parkinson	1.00	0.87	0.93	248
Accuracy			0.93	483
Macro Avg	0.94	0.93	0.93	483
Weighted Avg	0.94	0.93	0.93	483

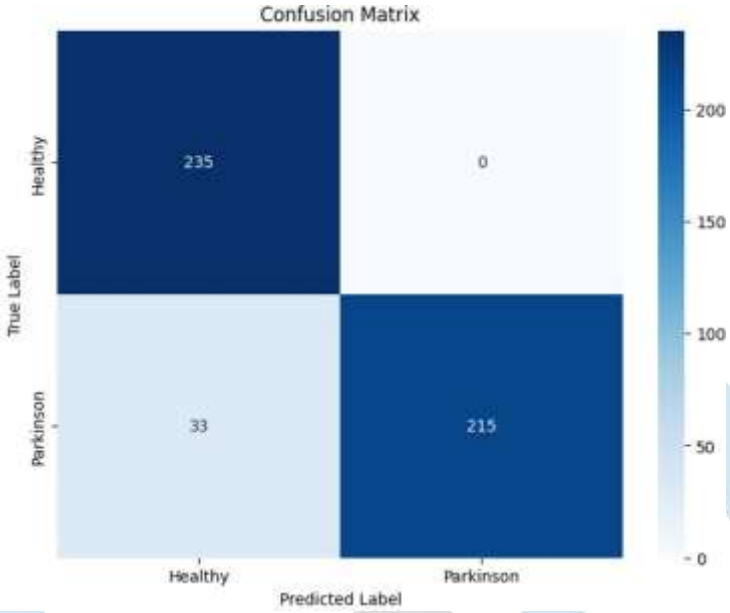


Fig. 4: Confusion matrix of the Parkinson's disease classification model using ResNet-18. All 250 healthy subjects were correctly classified, while 215 out of 248 Parkinson's cases were accurately identified. The matrix indicates high classification accuracy, with zero false positives and 33 false negatives.

By transitioning from Vision Transformers to ResNet-18, the model demonstrated an amazing improvement in classification performance with high accuracy and stable detection of Parkinson's disease via spiral images, as mentioned in Table I. ResNet-18's ability to extract hierarchical features effectively and robustness with relatively smaller datasets qualified it as the ideal candidate for this task. The combination of fine-tuned architecture, data augmentation, and training optimization created a highly reliable model that had the potential to assist clinicians in early intervention and diagnosis of PD. Future work will involve calibration with larger data, exploration of hybrid models that fuse CNNs and ViTs, and use of multimodal data to enhance the diagnostic capability of the model.

C. MRI Classification Performance (CNN)

The performance of the CNN-based classification model was evaluated using accuracy, pre- cision, recall, and F1-score metrics. The deep learning model achieved an overall classification accuracy of 95.33%, significantly outperforming the benchmark models. The performance metrics for all models are summarized in Table II:

The better performance of the CNN model can be explained by a number of important reasons:

1) **Hierarchical Feature Learning**: CNNs are specifically designed to learn spatial features from image data. The convolutional layers are capable of detecting local patterns like texture variations and local deformations in brain structures, which play an important role in differentiating PD from normal scans.

2) **Regularization**: Application of dropout layers at various points in the network, together with batch normalization, improves the generalizability of the model and reduces overfit- ting.
3) **End-to-End Optimization**: In contrast to RF and LR that rely on manually designed fea- tures, the CNN process does end-to-end joint feature learning and classification, resulting in task-specific and informative representations.

MRI imaging offers high-resolution structural data of subcortical brain structures, in particular the basal ganglia and substantia nigra, that are frequently impacted in Parkinson's Disease. As observed in Table II, the CNN model is able to well utilize this anatomical richness since it learns deep features that are not always obvious with visual analysis or traditional statistical methods.

Conversely, classical models like RF and LR, though effective for low-dimensional or tabular data, are not equipped to handle spatial hierarchies and contextual associations in images. This shortcoming accounts for their relatively poorer performance in the task of PD detection.

V. Conclusion

This research explored how deep learning can support early detection of Parkinson's Disease using two types of data: MRI brain scans and hand-drawn spiral images. By applying models like CNNs, LSTMs, and Vision Transformers, we were able to capture both structural brain changes and motor impairments. Among the models, ResNet-18 showed strong performance on

TABLE II: Performance comparison of different models on the MRI classification task

Model	Accuracy	Precision	Recall	F1-Score
CNN (Proposed)	95.33%	0.94	0.96	0.95
Random Forest (RF)	87.33%	0.86	0.89	0.87
Logistic Regression (LR)	85.33%	0.84	0.86	0.85

spiral image classification, showing that even with limited data, simple motor tasks can provide valuable insights.

The study highlights how AI can assist doctors by offering consistent and objective evaluations, especially when symptoms overlap or expert access is limited. These tools are not a replacement for clinical judgment but can strengthen decision-making in early and uncertain cases.

VI. Future Work

For future work, we aim to combine MRI and spiral data to build a more complete diagnostic system. Techniques like feature fusion and attention-based integration will help link motor function with brain structure for better accuracy. We're also working on making the system more accessible by developing a lightweight model for mobile screening using spiral drawings. In cases that require more detailed analysis, the system would recommend MRI follow-ups.

We also plan to study how well the model can track disease progression over time. The long- term vision is to create an AI-powered tool that is accurate, transparent, and usable in both well- equipped hospitals and remote clinics, helping neurologists make timely and informed decisions.

REFERENCES

[1] Beheshti, I. and Ko, J.H. (2024). Predicting the occurrence of mild cognitive impairment in Parkinson's disease using structural MRI data. *Frontiers in Neuroscience*, 18, p.1375395. https://doi.org/10.3389/fnins.2024.1375395

[2] Cao, K. et al. (2022). Identifying and validating subtypes of Parkinson's disease based on multimodal MRI data via hier- archical clustering analysis. *Frontiers in Human Neuroscience*, 16, p.919081. https://doi.org/10.3389/fnhum.2022.919081

[3] Chandra, J. et al. (2021). Screening of Parkinson's disease using geometric features extracted from spiral drawings. *Brain Sciences*, 11(10), p.1297. https://doi.org/10.3390/brainsci11101297

[4] Dadu, A. et al. (2022). Identification and prediction of Parkinson's disease subtypes and progression using machine learning in two cohorts. *NPJ Parkinson's Disease*, 8, p.172. https://doi.org/10.1038/s41531-022-00439-z

[5] DeSipio, R.E. (2023). Parkinson's Disease Automated Hand Tremor Analysis from Spiral Images.

[6] Depierreux, F. et al. (2021). Parkinson's disease multimodal imaging: F-DOPA PET, neuromelanin-sensitive and quantitative iron-sensitive MRI. *NPJ Parkinson's Disease*, 7, p.57. https://doi.org/10.1038/s41531-021-00199-2

[7] Elshewey, A.M. et al. (2023). Bayesian Optimization with Support Vector Machine Model for Parkinson Disease Classification. *Sensors*, 23, p.2085. https://doi.org/10.3390/s23042085

[8] Farhah, N. (2024). Utilizing deep learning models in an intelligent spiral drawing classification system for Parkinson's disease classification. *Frontiers in Medicine*, 11, p.1453743. https://doi.org/10.3389/fmed.2024.1453743

[9] Huang, Y. et al. (2024). Early Parkinson's Disease Diagnosis through Hand-Drawn Spiral and Wave Analysis Using Deep Learning Techniques. *Information*, 15, 220. https://doi.org/10.3390/info15040220

[10] Inguanzo, A. et al. (2024). MRI subtypes in Parkinson's disease across diverse populations and clustering approaches. *NPJ Parkinson's Disease*, 10, p.159. https://doi.org/10.1038/s41531-024-00759-2

[11] Jenei, A.Z., Sztaho´, D. and Vala´lik, I. Enhancing Parkinson's Disease Recognition through Multimodal Analysis of Archimedean Spiral Drawings.

[12] Lee, S.H. et al. (2022). Parkinson's disease subtyping using clinical features and biomarkers: Literature review and preliminary study of subtype clustering. *Diagnostics*, 12(1), p.112. https://doi.org/10.3390/diagnostics12010112

[13] Samantaray, T. et al. Labeling subtypes in a Parkinson's Cohort using Multifeatures in MRI: Integrating Grey and White Matter Information.

[14] Shakya, S. et al. (2022). Characterization of Parkinson's Disease Subtypes and Related Attributes. *Frontiers in Neurology*, 13, p.810038. https://doi.org/10.3389/fneur.2022.810038

[15] Sree Sai, A. et al. (2023). Early detection of Parkinson's Disease from spiral and wave drawings using image processing and machine learning techniques. *International Journal for Research Trends and Innovation*, 8(4), pp.58–61.

[16] Talai, A.S. et al. (2021). Utility of multi-modal MRI for differentiating Parkinson's disease and progressive supranuclear palsy using machine learning. *Frontiers in Neurology*, 12, p.648548. https://doi.org/10.3389/fneur.2021.648548

[17] Xiong, J. et al. (2023). Auto-Classification of Parkinson's Disease with Different Motor Subtypes Using Arterial Spin Labelling MRI Based on Machine Learning. *Brain Sciences*, 13, p.1524. https://doi.org/10.3390/brainsci13111524

[18] Yang, Y. et al. (2022). Identifying Depression in Parkinson's Disease by Using Combined Diffusion Tensor Imaging and Support Vector Machine. *Frontiers in Neurology*, 13:878691. https://doi.org/10.3389/fneur.2022.878691

[19] Yassine, S. et al. (2023). Identification of Parkinson's disease subtypes from resting-state electroencephalography. *Movement Disorders*, 38(8), pp.1451–1463. https://doi.org/10.1002/mds.29451

[20] Kingma, D.P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*. https: //arxiv.org/abs/1412.6980

[21] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(1), pp.1929–1958. http://jmlr.org/papers/ v15/srivastava14a.html

[22] Prechelt, L. (1998). Early Stopping—But When? In: Orr, G.B. and Mu¨ller, K.R. (eds) *Neural Networks: Tricks of the Trade*. Lecture Notes in Computer Science, vol 1524. Springer, Berlin, Heidelberg, pp.55–69. https://doi.org/10.1007/ 3-540-49430-8 3

[23] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), pp.5–32. https://doi.org/10.1023/A:1010933404324