

Article

An Uncertainty-Aware Transformer–Fuzzy Framework for Parkinson’s Disease Detection Using Handwritten Motor Patterns

Lipika Saluja ¹, Ayush Kumar Agrawal ¹, R Kanesaraj Ramasamy ^{2,*} and Parul Dubey ^{3,*}

¹ Department of Information Technology and Computer Science, Dr. C. V. Raman University, Bilaspur 495001, India; lipikadass03@gmail.com (L.S.); ayushagrawal@cvru.ac.in (A.K.A.)

² Faculty of Computing and Informatics, Multimedia University, Cyberjaya 63000, Malaysia

³ Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune 412115, India

* Correspondence: r.kanesaraj@mmu.edu.my (R.K.R.); parul.dubey@sitnagpur.siu.edu.in (P.D.)

Abstract

Parkinson’s disease is a progressive neurodegenerative disorder characterized by motor impairments that significantly affect handwriting and fine motor control. Recent advances in artificial intelligence have enabled the non-invasive analysis of handwritten patterns as reliable digital biomarkers for early Parkinson’s disease detection. However, existing deep-learning approaches often struggle with diagnostic uncertainty and lack interpretability, limiting their clinical reliability and practical adoption. Moreover, models trained on single datasets frequently exhibit poor generalization across heterogeneous handwriting sources. This study uses two image-based handwriting datasets and one CSV-based HandPD feature dataset, including the Parkinson’s Augmented Handwriting Dataset, Parkinson’s Drawings Dataset, and HandPD Spiral/Meander feature records. A Transformer-based architecture is employed to learn global motor patterns from handwriting images, followed by a fuzzy-logic-based decision layer to handle uncertainty and improve robustness. The novelty of this work lies in integrating Transformer-driven deep feature learning with fuzzy clinical reasoning, supported by an AIC-based handcrafted feature analysis for interpretability. The model performance is evaluated using accuracy, precision, recall, F1-score, MCC, and AUC metrics. The experimental results demonstrate that the proposed Transformer–Fuzzy framework consistently outperforms CNN and Transformer-only baselines, achieving superior classification performance and robust generalization across all datasets, thereby establishing its effectiveness for reliable and interpretable Parkinson’s disease screening.

Keywords: Parkinson’s disease; handwriting analysis; transformer networks; fuzzy logic; medical image classification

Academic Editor: Vincenzo Moscato

Received: 24 May 2026

Revised: 24 June 2026

Accepted: 24 June 2026

Published: 26 June 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Parkinson’s disease (PD) is a chronic progressive neurodegenerative disorder with more than 10 million cases all over the world, and prevalence estimates used to predict at least 1–2% of the population older than 60 years are affected [1]. By national estimates, the population of persons with PD may double by 2040 due to the aging of the U.S. population. Quantitatively, from a motor control standpoint, PD results in significant decrements

in fine motor performance as evidenced by increases in tremor amplitude, decreases in stroke smoothness, and increases in spatial variability [2,3]. For experimental handwriting studies, kinematic variability and trajectory irregularity are higher by approximately 30–60% for the PD patients over healthy controls with consistent statistical significance ($p < 0.01$) for both the spiral and wave drawing experiments [4,5]. These quantifiable differences define handwriting as a mathematically approachable and statistically significant biomarker for the early identification of Parkinson's disease [6].

Consequently, the handwriting-based analysis has become a low-cost and non-intrusive method to PD screening, especially for the early and prodromal stages [7]. Spiral- and wave-drawing tasks are frequently used in clinics, as they engage sustained motor coordination over time and are thus responsive to tremor, bradykinesia, and rigidity [8,9]. Recently, AI and deep-learning methods have been utilized for the automation of PD detection from handwritten motor tasks [10,11]. Convolutional Neural Networks (CNNs) have attained encouraging results in learning spatial features from handwriting images, but the utilization of local receptive fields will hinder the capability of modelling long-range dependencies, such as global stroke continuity and holistic motor coordination [12]. Transformer architectures based on self-attention mechanisms have shown promising performance in learning global contextual relations recently, which have also been effective for medical image analysis such as a handwriting-based motor assessment [13].

Despite this success, two issues remain unaddressed by the existing models for Parkinson's disease detection. First, the vast majority of deep-learning models (such as CNN and Transformer ones) are non-probabilistic and provide little information on the diagnostic uncertainty that is critical for clinical decision making [14,15]. Early-stage PD typically shows motor signs which are borderline or atypical, overlapping with elderly age symptoms and non-PD movement disorders; the binary distinction is too clean to be useful in real-life clinical practice [16,17]. Second, several current models tend to be black-box models focused on predictive performance, which restrains their interpretability and clinician reliance. As a result, there is a growing demand for intelligent diagnostic decision support systems that not only have a high detection sensitivity but also incorporate uncertainty and offer clinically relevant support in the decision-making process.

To overcome these constraints, the present work presents a hybrid Transformer–Fuzzy Logic architecture for the detection of Parkinson's disease by examining handwritten spiral and wave shapes. In the proposed method, a Transformer model is used to automatically capture long-range spatial dependencies and complex motor patterns, and learn discriminative global motor representations from images of handwriting based on self-attention mechanisms. Such deep features are fed to a Fuzzy Inference System (FIS), which is employed for the modelling of diagnostic uncertainty via linguistic variables, membership functions, and the expert-driven fuzzy rules.

This fusion facilitates uncertainty-aware decisions and interpretations by mapping continuous deep features to interpretable diagnostic scores, making the system closer to neurologist-minded reasoning. The contribution of the proposed work is in integrating distributed learning for clinical decision modelling with fuzzy logic so that the robust, interpretable, and uncertainty-aware detection of PD appears across the multiple handwriting datasets. The significant achievements of this study are summarized as follows:

- A hybrid Transformer–Fuzzy Logic model for Parkinson's disease identification, that leverages deep feature learning through a self-attention mechanism and an uncertainty-based approach for clinical decision modelling is presented;
- We design a handwriting analysis model based on the Transformer to address both global motor dependencies and fine-grained handwriting irregularities, which are not well-modelled in traditional CNN-based architectures;

- A Fuzzy Inference System is included as a post-decision layer for addressing the diagnostic uncertainty, borderline situation, and gradual progression of symptoms, thus improving the clinical interpretability.
- The performance of the proposed methodology is verified on several public handwriting datasets with the focus on robustness and cross-dataset generalization rather than single-dataset maximization;
- The study presents a potentially useful and interpretable AI-assisted screening framework that may support future Parkinson's disease decision-support studies, subject to further validation on prospective clinical data.

As illustrated in Figure 1, the end-to-end pipeline of our proposed approach consists of the following: multi-scene handwriting inputs for fine-grained motor impairments; Transformer-based global feature learning (m-transformer); and fuzzy-logic-driven clinical reasoning to handle uncertainty towards confident diagnostic decisions and practical clinical decision support.

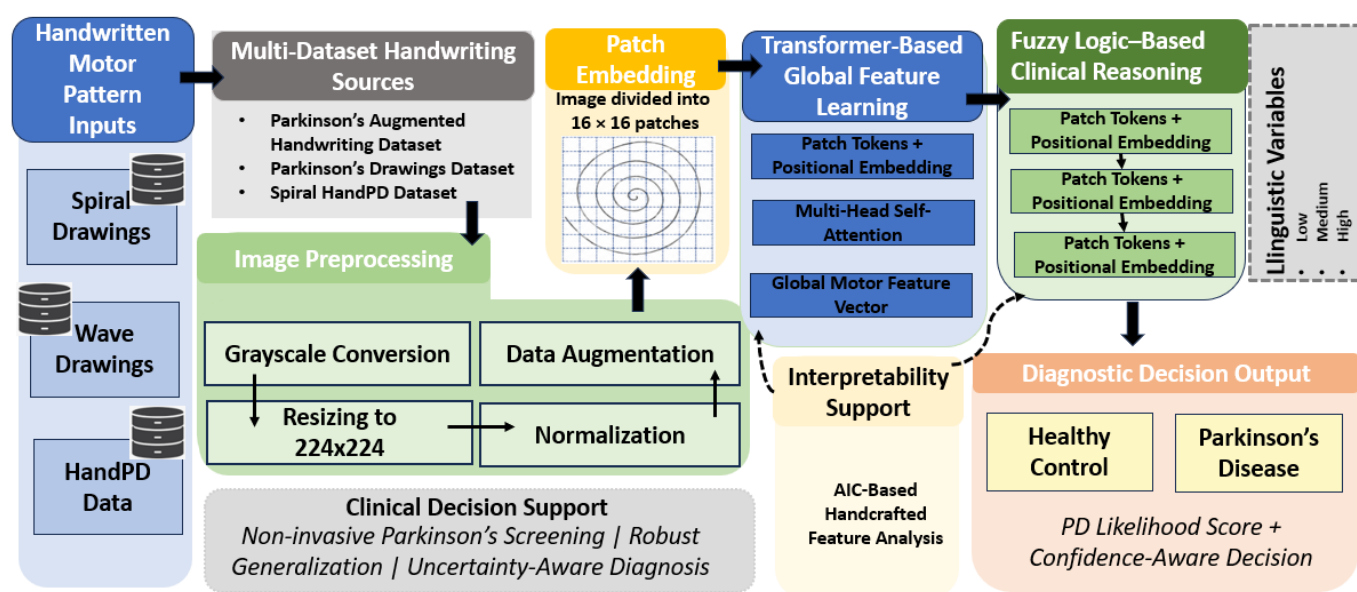


Figure 1. Overview of the proposed Transformer–Fuzzy Framework for Parkinson’s disease detection.

2. Literature Review

Parkinson’s disease (PD) is a devastating, neurodegenerative disorder with the combination of motor and non-motor symptoms significantly impairing the quality of life in patients. Clinical trials have also emphasized the role of valid biomarkers for early diagnosis and to help in distinguishing among diseases, particularly to differentiate PD from atypical parkinsonian syndromes, and another neurological mimicker [18]. Stronger correlations, between striatal DAT binding and neurotransmitter system derangement as well as motor or nonmotor symptom severity, were found in early PD patients using neuroimaging or biochemical measurements [19,20]. Several composite clinical scales and symptom prevalence studies also underscore the heterogeneous nature of PD progression, and the need for objective, quantification assessment measures [21].

In the past few years, handwriting analysis has been proposed as a potential digital non-invasive PD biomarker. The first works have investigated dynamic writing traces, and signal warping methods for encoding tremor, dyskinesia, and deviations in handwriting trajectories [22]. Early signs of aberrant handwriting, such as micrographic, stroke distortion, and poor motor fluency, have been identified in PD [23]. These results

demonstrated the clinical utility of spiral- and wave-drawing tasks as performance-based measures.

The automatic diagnosis of PD using handwriting has gained significant interest due to the recent advances in machine learning. Some previous studies explored classical machine learning (ML), and shallow neural network (NN) models for PD detection by means of the handcrafted handwriting features [8,15,16]. Subsequently, deep-learning-based techniques such as RNNs and CNNs achieved better performance by learning complex temporal and spatial features from online/offline handwritten data [24]. Multimodal systems further improved diagnostic accuracy by adding handwriting to other physiological signals such as EEG but at the price of a commensurate increase in complexity.

Finally, deep-transfer-learning and attention-based models are increasingly popular for improving generalization performance and robustness. As far as the PD classification on handwritings, several works employed pre-trained CNNs and training techniques to gain a significant accuracy improvement [25,26]. Transparent and clinically interpretable AI models have also been developed, especially with the help of explainable AI (XAI) techniques which advance model interpretability feature attribution and visualization methods. Ensemble learning has also been used for superior performance by combining multimodal feature representations from different networks [27].

Transformer-based models are the recent inclusions to handwriting analysis: they enable the modelling of long-range dependencies and global context. These attention-mechanism-based models, such as the Swin Transformer and coordinate-based attention mechanism, were also feasible in handwriting recognition and PD detection [26]. However, the Transformer models are not equipped with explicit clinical interpretation and uncertainty estimation to handle diagnostic uncertainty specifically, which causes unstable predictions for cases located near the borderline, which is essential for clinical reliability.

The original paper of Pradeep et al. also proposed an improved LinkNet–GhostNet architecture for Parkinson’s disease detection in handwriting images and achieved a strong classification performance on the spiral- and wave-drawing datasets [25]. The work demonstrated well the promise of employing lightweight deep models for PD screening, and set a strong benchmark in handwriting-based detection. But this model was mostly built upon deterministic deep learning prediction with less attention on uncertainty processing and decision explainability. Furthermore, the method placed emphasis on convolutional feature extraction, which might hinder the modelling of global stroke dependencies existing in handwriting patterns.

The existing work makes it clear that, although deep-learning- and Transformer-based methods have advanced the performance of handwriting-based PD detection, challenges such as interpretability, and uncertainty modelling across datasets are not well addressed. The majority of studies are concentrated on performance improvement or explainability alone and there is scarcely any attempt to jointly integrate both into a single framework. This gap is what motivates the current study, which combines Transformer-based deep feature learning with fuzzy-logic-based decision modelling to increase robustness, interpretability, and clinical reliability across different handwriting datasets. Table 1 summarizes the literature review.

Table 1. Summary of key literature on Parkinson’s disease detection using handwriting and AI-based methods.

Reference(s)	Focus/Modality	Method Used	Key Contribution	Research Gap
Deharab and Ghaderyan [1,7]; Senatore et al. [4]	Handwriting traces and early PD signs	Signal analysis and handwriting variability assessment	Established handwriting irregularity, tremor, and trajectory variation as useful PD indicators	Limited deep learning and uncertainty-aware modelling
Zhang et al. [3]; Kumar and Ghosh [6]; Arasavali et al. [9]; Randazzo et al. [10]	Handwriting-based PD detection	Deep learning, RNN, MLP, and neural feature extraction	Demonstrated effectiveness of neural models for handwriting motor-pattern learning	Limited interpretability and fuzzy clinical reasoning
Thebaud et al. [5]; Murugaveni et al. [22]	Interpretable handwriting analysis	Handcrafted features and explainable AI	Improved clinical interpretability of handwriting-based PD assessment	Limited fuzzy uncertainty modelling
Kumar and Ghosh [8]; Aljohani [20]	Multimodal/feature-fusion PD detection	EEG–handwriting fusion and voting classifier	Showed that combining features/signals can improve diagnostic performance	Increased complexity and lower practical simplicity
Alsalem and Amin [18]; Shanmugavadivu et al. [19]; Abdullah et al. [21]	Image-based PD detection	Transfer learning and optimized feature selection	Improved image-based classification performance	Mostly deterministic; limited uncertainty handling
Al-Jabbar et al. [24]	Spiral and wave handwriting drawings	Multi-CNN and ensemble learning	Provided strong recent benchmark using hybrid CNN and ensemble models	Complex architecture; limited fuzzy interpretability
Pradeep and Kamalakannan [25]	Handwriting image-based PD detection	Improved LinkNet–GhostNet	Achieved strong performance using lightweight deep architecture	Limited rule-based and uncertainty-aware decision support
Wang et al. [26]	Handwriting recognition for PD	Coordinate attention-enhanced Swin Transformer	Demonstrated usefulness of Transformer-based attention for handwriting analysis	No direct fuzzy inference or uncertainty modelling
Public datasets [28–30]	Spiral, wave, and HandPD handwriting data	Public handwriting datasets	Provide benchmark data for PD and healthy control classification	Dataset size, subject metadata, and leakage control remain concerns

3. Dataset Description

This study uses two image-based public handwriting datasets and one CSV-based HandPD feature dataset. The image datasets were used for Transformer-based visual feature learning, whereas the HandPD CSV files were used for feature-level motor-pattern analysis and fuzzy reasoning support. Handwriting tasks have been established as clinically valid indicators for Parkinsonian motor deformation, as they measure the tremor amplitude, stroke non-regularity and fine motor coordination degradation. The chosen datasets have different data acquisition conditions, drawing styles, and sample distributions to avoid overfitting on a specific dataset.

3.1. Parkinson’s Augmented Handwriting Dataset

Parkinson’s Augmented Handwriting Dataset is a dataset of handwritten spiral and wave gestures from patients with Parkinson’s disease [28]. The dataset was specifically designed to address the lack of data via augmentation approaches such as rotation, scaling, translation, and flipping. These alterations model the natural variation in handwriting

and preserve Parkinson's disease associated motor characteristics. The combination of spiral and wave tasks allows the proposed motor patterns, such as tremor severity, directional inconsistency, or spatial distortion, to be learned by the model. The presence of data augmentation in this set is especially interesting for the training of larger transformer models, which require large and varied sample distributions to achieve an effective self-attention learning process.

3.2. Parkinson's Drawings Dataset

The Parkinson's Drawings Dataset consists of static spiral and wave drawings drawn by patients with Parkinson's disease and normal subjects [29]. This dataset is different from the augmented datasets, since it mostly includes raw and unprocessed images of handwriting which reflects the true variability in real drawing as well as acquisition noise. The dataset is generally partitioned into training and test sets which make the experimental evaluation universally applicable. This motor phenomena in Parkinson's disease patients can be easily observed visibly through the tremor-corrupted shapes, varying thickness of strokes, and smoothening of drawings. This dataset is a crucial benchmark for assessing the generalization of the model for non-augmented, clinically relevant handwriting data.

3.3. HandPD CSV Dataset

The HandPD dataset used in this study was obtained in CSV format rather than as raw handwriting images. Therefore, it was treated as a feature-level handwriting dataset. Two files were used: Spiral_HandPD.csv and Meander_HandPD.csv [30]. Each file contains 368 records, 16 variables, and 92 unique examination/session IDs. The class label is provided as CLASS_TYPE, where Class 1 represents Healthy Control and Class 2 represents Parkinson's Disease. In both files, 72 records belong to Healthy Controls and 296 records belong to Parkinson's Disease cases, corresponding to 18 healthy and 74 Parkinson's examination/session IDs.

The dataset includes metadata such as examination ID, image reference, patient ID, gender, handedness, and age, along with numerical handwriting descriptors such as RMS, maximum- and minimum-height-related measures, standard deviation measures, mean response/time-related measures, and changes from negative to positive movement patterns. Since raw handwriting images were not available, HandPD was not subjected to image resizing, patch extraction, or image-based preprocessing. Instead, numerical feature normalization was applied, and the dataset was used to support feature-level motor-pattern analysis and fuzzy reasoning.

For the HandPD CSV data, splitting was performed at the examination/session level using _ID_EXAM as the grouping variable to avoid leakage from the same examination/session appearing in more than one subset. Strict subject-wise separation could not be guaranteed for datasets where participant identifiers were unavailable. The overview of the handwriting datasets we employed in this work is given in Table 2.

Table 2. Dataset summary, data format, class distribution, and splitting strategy.

Dataset	Task Type and Data Format	Total Samples Used	Class Distribution	Subject/Exam Information	Split Used in This Study	Leakage-Control Strategy
Parkinson's Spiral and Augmented Handwriting Dataset	Spiral and wave handwriting images	3264 images	1632 Healthy Control; 1632 Parkinson's Disease	Subject IDs not consistently available	70:15:15 split; Train = 2284, Validation = 490, Test = 490	Augmented variants derived from the same original image were kept within the same subset

Parkinson's Drawings Dataset	Spiral and wave drawing images	204 images	102 Healthy Control; 102 Parkinson's Disease	27 subjects reported in dataset description; subject IDs not available in folder-level metadata	70:15:15 split; Train = 142, Validation = 31, Test = 31	Subject-wise separation was applied where subject identifiers were available; otherwise, stratified image-level splitting was used
HandPD Spiral CSV and HandPD Meander CSV	CSV feature records	368 records	72 Healthy Control records; 296 Parkinson's Disease records	92 exam IDs; 18 Healthy Control exam IDs and 74 Parkinson's Disease exam IDs	Grouped 70:15:15 split using _ID_EXAM; Train = 260, Validation = 52, Test = 56	All records from the same _ID_EXAM were kept in only one subset

4. Experimental Setup

All experiments were conducted on a server with an Intel Core i7 CPU, 16 G memory, and a NVIDIA GPU with CUDA support to speed up deep-learning calculations. The proposed framework was implemented using Python 3.10 and PyTorch 2.0 as the main deep-learning library for realizing the Transformer tool. Other libraries including NumPy 1.24, OpenCV 4.8, and Scikit-learn 1.3 have been used for image preprocessing, feature manipulation, and performance assessment. The fuzzy inference module utilized a home-built rule-based fuzzy logic layer for flexibility and interpretability.

All handwritten images taken from the selected datasets were converted to grayscale and ensured to have a regular resolution of 224×224 pixels, so as to align all the photographs across datasets. This image preprocessing was applied only to the image-based datasets. For the HandPD CSV dataset, numerical feature normalization was applied instead. Pixel values were all scaled between $[0, 1]$ to increase the robustness of gradient-based optimization. Light data augmentation including random rotation, translation, and scaling was further used for training to enhance generalization and alleviate overfitting. No augmentation was used during validation and test.

The proposed model uses a Vision Transformer (ViT)-based encoding to capture global motor representations from handwriting images. We split each input image into 16×16 non-overlapping patches, which were then flattened and linearly transformed to a latent feature space. Positional embeddings were used to keep the spatial information. The Transformer encoder is composed of stacked layers of multi-head self-attention and feed-forward networks, which gives the model the capability to capture long-range spatial dependencies including stroke continuity, tremor propagation, and global motor coordination patterns.

The Transformer was trained with an Adam optimizer and a fixed learning-rate, stopping early to avoid overfitting based on validation loss. Uncertainty-aware clinical decision making was injected into the Transformer-extracted features as a post-decision layer using the Fuzzy Inference System (FIS). The extracted deep features were then converted to linguistic variables, representing the motor impairment intensity (e.g., Low/Medium/HIGH). Triangular membership functions have been employed for fuzzification because of its simplicity and interpretability. To simulate incremental or borderline progression, a rule set based on expert opinion that utilized fuzzy IF–THEN rules was developed. The last diagnostic conclusions were achieved via centroid defuzzification and a threshold was employed to produce binary classification outputs.

4.1. Data Splitting, Leakage Control, and Evaluation Protocol

A structured training, validation, and testing protocol was adopted to improve reproducibility and reduce the risk of optimistic performance estimation. For the image-based datasets, namely, the Parkinson's Augmented Handwriting Dataset and Parkinson's Drawings Dataset, images were divided into training, validation, and testing subsets using a 70:15:15 stratified split. Stratification was applied to preserve the Healthy Control and Parkinson's Disease class distribution across all subsets.

For the HandPD Spiral and Meander CSV files, splitting was performed at the examination/session level using `_ID_EXAM` as the grouping variable. This ensured that all records belonging to the same examination/session were assigned to only one subset. Under the grouped 70:15:15 split, each HandPD CSV file was divided into 260 training records, 52 validation records, and 56 testing records.

For the augmented handwriting dataset, augmented samples were grouped with their corresponding original images only when the original–augmented relationship could be verified from filename or metadata information. Because this relationship could not be confirmed for all samples, the possibility of residual augmentation-related leakage cannot be completely ruled out and is acknowledged as a limitation.

Three evaluation settings were used. First, within-dataset evaluation was performed separately on each dataset to assess dataset-specific performance. Second, a combined-dataset evaluation was conducted by pooling compatible samples after applying the same preprocessing and normalization procedures. Third, cross-dataset validation was performed to examine whether the model can be generalized across heterogeneous handwriting sources. In cross-dataset validation, the model was trained on one or more datasets and tested on a separate dataset that was not used during training. This setting was used to assess external generalization rather than only random-split performance.

All experiments were repeated across multiple independent runs using fixed random seeds. The reported results were averaged across runs, and the variability was summarized using standard deviation and confidence intervals wherever applicable.

4.2. Performance Evaluation Metrics

Standard healthcare diagnostics metrics such as Accuracy, Precision, Recall (sensitivity), F1-score, Matthews Correlation Coefficient (MCC), Area Under the ROC Curve (AUC), and Specificity were used for evaluating the behaviour of the Transformer–Fuzzy system. In combination, the three measures assess classification certainty, performance robustness, and relevance to clinical considerations. The experimental setting and training options for the development and evaluation of our models are summarized in Table 3.

Table 3. Experimental setup and training configuration.

Component	Parameter	Value/Setting
Programming environment	Programming language	Python 3.x
Deep-learning framework	Framework	PyTorch
Hardware	System configuration	Intel Core i7 CPU, 16 GB RAM, NVIDIA CUDA-supported GPU
Image preprocessing	Image input size	224 × 224 pixels for image-based datasets only
	Image type	Grayscale
	Pixel normalization	Min–Max scaling to [0, 1]
CSV preprocessing	HandPD CSV features	Numerical features normalized to [0, 1]
Data augmentation	Augmentation strategy	Random rotation, translation, and scaling applied only to training images
Patch embedding	Patch size	16 × 16 for image-based Transformer input
Transformer encoder	Embedding dimension	768
	Number of encoder layers	12
	Attention heads	12

	Dropout	0.1
Optimizer	Optimizer type	Adam
Optimization	Learning rate	0.0001
	Batch size	32
Training	Maximum epochs	50
	Early-stopping criterion	Validation loss
	Early-stopping patience	10 epochs
Loss function	Classification loss	Binary cross-entropy
Reproducibility	Random seeds	42, 123, 2024, 2025, 2026
	Repeated runs	5 independent runs
Data split	Split ratio	70:15:15 train–validation–test split
	HandPD split strategy	Grouped split using _ID_EXAM
Validation strategy	Evaluation settings	Within-dataset, combined-dataset, and cross-dataset evaluation
	Fuzzy input variables	Tremor severity, stroke irregularity, and global distortion
Fuzzy inference	Membership function	Triangular membership functions
	Linguistic levels	Low, Medium, and High
	Rule base	27 Mamdani IF–THEN rules
	Defuzzification method	Centroid defuzzification
	Decision threshold	0.5
Evaluation metrics	Classification metrics	Accuracy, Precision, Recall/Sensitivity, Specificity, F1-score, MCC, and AUC
	Uncertainty metrics	Brier score and Expected Calibration Error (ECE)

5. Proposed Methodology

In this section, the proposed Transformer–Fuzzy Logic hybrid model for the detection of Parkinson’s disease using handwritten motor tasks is presented. The methodology combines deep self-attention-based representation learning and fuzzy rule-based inference for uncertainty-aware and clinically interpretable decision-making. Figure 2 shows the internal structure of our system, in which handwriting images are initially represented by fixed-size patches and encoded with a Vision Transformer to capture global motor characteristics. The resulting feature vector is then fed into a fuzzy inference system consisting of fuzzification, rule-based reasoning and centroid-based defuzzification to generate an uncertainty-aware diagnostic decision as well as a clinically interpretable confidence score.

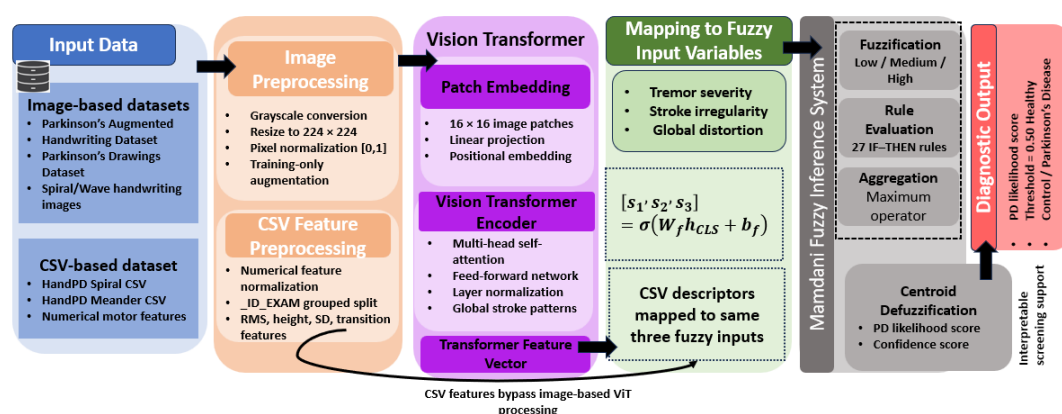


Figure 2. Detailed architecture of the proposed Transformer–Fuzzy Inference framework for Parkinson’s disease detection.

5.1. Problem Formulation

Equation (1) denotes the dataset:

$$\mathcal{D} = \{(I_i, y_i)\}_{i=1}^N \quad (1)$$

where $I_i \in \mathbb{R}^{H \times W}$ represents a handwriting image (spiral or wave) and $y_i \in \{0,1\}$ denotes the ground-truth class label (0: Healthy, 1: Parkinson's Disease).

The objective is to learn a mapping, as shown in Equation (2),

$$f: I \rightarrow \hat{y} \quad (2)$$

that maximizes classification performance while explicitly accounting for diagnostic uncertainty.

The proposed pipeline is expressed as per Equation (3):

$$I \xrightarrow{\text{Transformer}} F \xrightarrow{\text{Fuzzy Logic}} \hat{y} \quad (3)$$

5.2. Transformer-Based Feature Learning

The Transformer module is responsible for learning global motor representations from handwriting images by modelling long-range spatial dependencies.

5.2.1. Patch Embedding

Each preprocessed image $I \in \mathbb{R}^{224 \times 224}$ is divided into non-overlapping patches of size $P \times P$. Equation (4) shows total number of patches:

$$N_p = \frac{HW}{P^2} \quad (4)$$

Each patch p_i is flattened and projected into a D -dimensional latent space, as per Equation (5):

$$x_i = p_i W_E + b_E \quad (5)$$

A learnable class token x_{cls} and positional embeddings E_{pos} are added, as shown in Equation (6):

$$X = [x_{cls}; x_1; x_2; \dots; x_{N_p}] + E_{pos} \quad (6)$$

5.2.2. Self-Attention Mechanism

For an input sequence X , the query, key, and value matrices are computed as per Equation (7):

$$Q = XW_Q, K = XW_K, V = XW_V \quad (7)$$

Equation (8) shows the scaled dot-product attention:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (8)$$

Multi-head attention aggregates information from multiple subspaces, as per Equation (9):

$$\text{MHSA}(X) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O \quad (9)$$

This enables the model to capture global handwriting characteristics, such as tremor propagation and stroke continuity.

5.2.3. Transformer Encoder Output

Each Transformer layer consists of MHSA and a feed-forward network (FFN), as shown in Equation (10):

$$\text{FFN}(x) = \sigma(xW_1 + b_1)W_2 + b_2 \quad (10)$$

Residual connections and layer normalization are applied as per Equation (11):

$$\begin{aligned} Z^{(l)} &= \text{LN}(\text{MHSA}(Z^{(l-1)}) + Z^{(l-1)}) \\ Z^{(l)} &= \text{LN}(\text{FFN}(Z^{(l)}) + Z^{(l)}) \end{aligned} \quad (11)$$

The final class token representation z_{cls} is extracted as the global feature vector, shown in Equation (12):

$$F = z_{\text{cls}} \quad (12)$$

5.3. Fuzzy-Logic-Based Clinical Decision Modelling

Even though Transformer features are highly discriminatory, the clinical diagnosis is of gradual development and not certain. For this, a Fuzzy Inference System (FIS) is used. Figure 3 shows the fuzzy inference performed on normalized motor features obtained from the Transformer model. Tremor intensity, stroke irregularity, and global distortion transformer outputs are initially fuzzified into linguistic terms (Low, Medium, and High), and then the Mamdani inference rule evaluation is applied along with aggregation. The last step is by centroid-based defuzzification leading to a Parkinson's disease likelihood score, which provides the uncertainty-aware and clinically interpretable diagnostic decision.

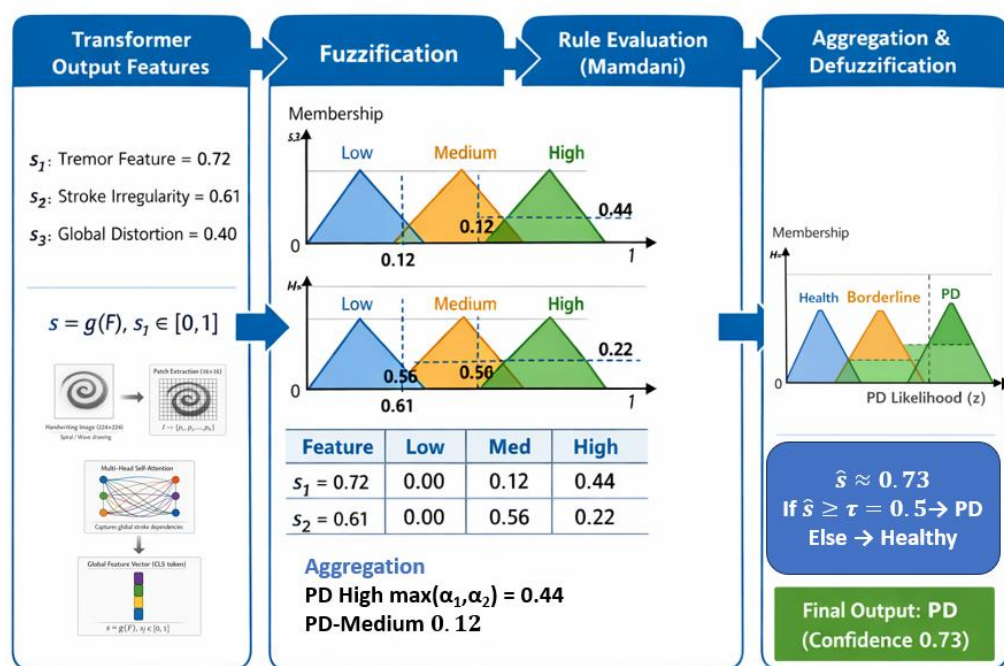


Figure 3. Mamdani fuzzy inference process showing fuzzification of normalized motor-impairment scores, rule firing, aggregation, and centroid-based PD likelihood estimation.

For example, if the normalized fuzzy input scores are tremor severity = 0.72, stroke irregularity = 0.61, and global distortion = 0.40, each value is first converted into Low, Medium, and High membership degrees using the triangular membership functions. Suppose tremor has a High membership of 0.30 and stroke irregularity has a High membership of 0.05. For the rule “IF tremor severity is High AND stroke irregularity is High THEN PD likelihood is High,” the Mamdani firing strength is calculated using the minimum operator: $\min(0.30, 0.05) = 0.05$. Similarly, all activated rules are evaluated, their outputs are aggregated using the maximum operator, and the final PD likelihood score is obtained using centroid defuzzification. A score equal to or above 0.50 is classified as Parkinson's Disease, while a score below 0.50 is classified as Healthy Control.

5.3.1. Mapping of Features to Fuzzy Input Variables

The Fuzzy Inference System requires interpretable scalar inputs instead of the complete high-dimensional Transformer feature vector. Therefore, the final Transformer class-token representation h_{CLS} was projected into three normalized motor-impairment scores: tremor severity, stroke irregularity, and global distortion. This mapping was performed using a linear projection layer followed by sigmoid normalization shown in Equation (13):

$$[s_1, s_2, s_3] = \sigma(W_f h_{CLS} + b_f) \quad (13)$$

where s_1 , s_2 , and s_3 represent tremor severity, stroke irregularity, and global distortion, respectively. Each score ranges from 0 to 1, with higher values indicating stronger hand-writing impairment.

For the HandPD CSV dataset, raw images were not available; therefore, normalized numerical descriptors were grouped into the same three fuzzy inputs. RMS and height-variation features were used for tremor severity, standard deviation, and transition-related features for stroke irregularity, and maximum-/minimum-height-related features for global distortion. These three inputs were then fuzzified into Low, Medium, and High linguistic terms. To ensure a consistent fuzzy reasoning process across image-based and CSV-based datasets, the extracted features were mapped into three common fuzzy input variables, as shown in Table 4.

Table 4. Mapping of image-based and HandPD CSV features to fuzzy input variables.

Fuzzy Input	Image-Based Datasets	HandPD CSV Dataset
Tremor severity	Projected Transformer feature	RMS and height-variation features
Stroke irregularity	Projected Transformer feature	Standard deviation and transition features
Global distortion	Projected Transformer feature	Maximum-/minimum-height features

5.3.2. Fuzzification and Membership Parameters

Selected Transformer features $F = \{f_1, f_2, \dots, f_m\}$ are mapped to linguistic variables such as Low, Medium, and High motor impairment.

Equation (14) defines the Triangular membership functions:

$$\mu_A(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ \frac{c-x}{c-b}, & b < x \leq c \\ 0, & x > c \end{cases} \quad (14)$$

The three fuzzy input variables, namely, tremor severity, stroke irregularity, and global distortion, were normalized within the range $[0, 1]$. Each input variable was represented using three triangular membership functions: Low, Medium, and High. The same membership parameters were used for all three input variables to maintain consistency and interpretability. The triangular membership parameters used for the fuzzification of the three motor-impairment inputs and the PD likelihood output are presented in Table 5.

Table 5. Triangular membership function parameters for fuzzy inputs and output.

Variable	Low	Medium	High
Tremor severity	(0.00, 0.00, 0.40)	(0.20, 0.50, 0.80)	(0.60, 1.00, 1.00)
Stroke irregularity	(0.00, 0.00, 0.40)	(0.20, 0.50, 0.80)	(0.60, 1.00, 1.00)
Global distortion	(0.00, 0.00, 0.40)	(0.20, 0.50, 0.80)	(0.60, 1.00, 1.00)
PD likelihood output	Healthy: (0.00, 0.00, 0.40)	Borderline: (0.25, 0.50, 0.75)	PD: (0.60, 1.00, 1.00)

5.3.3. Fuzzy Rule Base

A set of expert-inspired IF–THEN rules is constructed. The following is an example:

IF tremor severity is High AND stroke irregularity is High, THEN PD likelihood is High.

The firing strength of the k -th rule is computed using the min operator, as per Equation (15):

$$\alpha_k = \min(\mu_{A_1}(f_1), \mu_{A_2}(f_2), \dots) \quad (15)$$

A total of 27 Mamdani IF–THEN rules were used because the system contains three input variables, each with three linguistic levels. To keep the manuscript concise, we can describe the rule base using this compact Table 6.

Table 6. Compact Mamdani fuzzy-rule-base design for diagnostic decision-making.

Condition Pattern	Output Decision
All three inputs Low	Healthy
Two inputs Low and one Medium	Healthy
One input High with remaining inputs Low/Medium	Borderline
Two inputs Medium and one Low/Medium	Borderline
Tremor High + Stroke High	Parkinson's Disease
Tremor High + Global distortion High	Parkinson's Disease
Stroke High + Global distortion High	Parkinson's Disease
All three inputs High	Parkinson's Disease

The rules were designed using clinically motivated reasoning. Low values across all motor-impairment variables indicate healthy handwriting patterns, whereas high tremor, high stroke irregularity, and high global distortion indicate Parkinsonian handwriting impairment. Borderline outputs were assigned when the variables showed mixed or moderate impairment patterns. The rule firing strength was computed using the Mamdani minimum operator, and all activated rule outputs were aggregated using the maximum operator.

5.3.4. Defuzzification

The aggregated fuzzy output is converted into a crisp diagnostic score using the centroid method, as per Equation (16):

$$\hat{y} = \frac{\int z \cdot \mu(z) dz}{\int \mu(z) dz} \quad (16)$$

A threshold τ is applied for binary classification is shown in Equation (17):

$$\hat{y} = \begin{cases} 1, & \hat{y} \geq \tau \\ 0, & \hat{y} < \tau \end{cases} \quad (17)$$

The approach is the first to couple Transformer-based global motor representation learning with fuzzy-logic-driven uncertainty modelling, and, thus, it allows the transparent and clinically robust detection of Parkinson's disease that goes beyond deterministic

deep-learning classifiers. The aggregated fuzzy output was converted into a crisp PD likelihood score using centroid defuzzification. A final threshold of 0.50 was applied, where scores below 0.50 were classified as Healthy Control and scores equal to or above 0.50 were classified as Parkinson's Disease.

The AIC-based handcrafted feature analysis was used only as a post hoc interpretability component and was not used as a direct input to the Transformer classifier. The proposed Transformer–Fuzzy model performs the classification using Transformer-derived representations and fuzzy inference. In contrast, the AIC analysis was applied separately to identify compact and clinically meaningful handwriting descriptors associated with Parkinsonian motor impairment. Therefore, the AIC component supports the interpretation of handwriting-related motor features but does not directly influence the model training or final prediction. Algorithm 1 below explains the process mathematically.

Algorithm 1: Unified Transformer–Fuzzy Logic Method (Mathematical Algorithm)

Inputs

- Dataset: $\mathcal{D} = \{(I_i, y_i)\}_{i=1}^N$, $I_i \in \mathbb{R}^{H \times W}$, $y_i \in \{0,1\}$
- Patch size P , embedding dim D , Transformer depth L , heads h
- Fuzzy linguistic sets $\mathcal{L} = \{\text{Low}, \text{Med}, \text{High}\}$
- Membership parameters $\Theta = \{(a_{j\ell}, b_{j\ell}, c_{j\ell})\}$ for feature j and label ℓ
- Rule base $\mathcal{R} = \{r_k\}_{k=1}^K$
- Decision threshold $\tau \in (0,1)$

Step 1: Preprocessing

For each image I , define the standardized input

$$\tilde{I} = \phi(I) \in \mathbb{R}^{224 \times 224}, \tilde{I} = \frac{\text{ResizeGray}(I) - \min}{\max - \min}.$$

Step 2: Patchification and Embedding

Partition $\tilde{I}$ into $N_p = \frac{HW}{P^2}$ non-overlapping patches $\{p_n\}_{n=1}^{N_p}$, $p_n \in \mathbb{R}^{P \times P}$. Flatten and embed each patch:

$$x_n = \text{vec}(p_n)W_E + b_E \in \mathbb{R}^D.$$

Add a learnable class token $x_{\text{cls}} \in \mathbb{R}^D$ and positional encodings E_{pos} :

$$X_0 = [x_{\text{cls}}; x_1; \dots; x_{N_p}] + E_{\text{pos}} \in \mathbb{R}^{(N_p+1) \times D}.$$

Step 3: Transformer Feature Learning

For layer $\ell = 1, \dots, L$, compute multi-head self-attention and FFN updates:

(a) MHSA:

$$\begin{aligned} Q &= X_{\ell-1}W_Q, K = X_{\ell-1}W_K, V = X_{\ell-1}W_V, \\ \text{Att}(Q, K, V) &= \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V, \\ \text{MHSA}(X_{\ell-1}) &= \text{Concat}(\text{Att}_1, \dots, \text{Att}_h)W_O. \end{aligned}$$

(b) Residual + Norm, then FFN:

$$\begin{aligned} \hat{X}_\ell &= \text{LN}(X_{\ell-1} + \text{MHSA}(X_{\ell-1})), \\ X_\ell &= \text{LN}(\hat{X}_\ell + \text{FFN}(\hat{X}_\ell)), \text{FFN}(u) = \sigma(uW_1 + b_1)W_2 + b_2. \end{aligned}$$

Finally, extract the global representation:

$$F = X_L[0] \in \mathbb{R}^D (\text{class-token embedding}).$$

Step 4: Feature Selection / Projection for Fuzzy Reasoning

Select m fuzzy-relevant features via a projection $g(\cdot)$:

$$s = g(F) \in \mathbb{R}^m.$$

(Example: g can be a linear projection $s = WF + b$, or a selection of specific components.)

Step 5: Fuzzification (Triangular Membership)

For each fuzzy input feature s_j and linguistic label $\ell \in \mathcal{L}$, define triangular membership:

$$\mu_{j\ell}(s_j) = \begin{cases} 0, & s_j \leq a_{j\ell}, \\ \frac{s_j - a_{j\ell}}{b_{j\ell} - a_{j\ell}}, & a_{j\ell} < s_j \leq b_{j\ell}, \\ \frac{c_{j\ell} - s_j}{c_{j\ell} - b_{j\ell}}, & b_{j\ell} < s_j < c_{j\ell}, \\ 0, & s_j \geq c_{j\ell}. \end{cases}$$

Step 6: Rule Evaluation (Mamdani Inference)

Each rule r_k has the form:

$$r_k: \text{IF } (s_1 \text{ is } A_{k1}) \wedge \dots \wedge (s_m \text{ is } A_{km}) \text{ THEN } z \text{ is } B_k,$$

where $A_{kj} \in \mathcal{L}$ and B_k is an output fuzzy set over $z \in [0,1]$ ("PD likelihood").

Rule firing strength using the minimum t-norm:

$$\alpha_k(s) = \min_{j=1,\dots,m} \mu_{j,A_{kj}}(s_j).$$

Implication (Mamdani "clipping"):

$$\mu'_{B_k}(z) = \min(\alpha_k(s), \mu_{B_k}(z)).$$

Aggregation across rules (maximum s-norm):

$$\mu_{\text{agg}}(z) = \max_{k=1,\dots,K} \mu'_{B_k}(z).$$

Step 7: Defuzzification and Decision

Centroid defuzzification yields a crisp PD score $\hat{s} \in [0,1]$:

$$\hat{s} = \frac{\int_0^1 z \mu_{\text{agg}}(z) dz}{\int_0^1 \mu_{\text{agg}}(z) dz}.$$

Final predicted label:

$$\hat{y} = \begin{cases} 1, & \hat{s} \geq \tau, \\ 0, & \hat{s} < \tau. \end{cases}$$

Training Objective (Transformer parameters)

Let Transformer parameters be Ω and fuzzy parameters (membership + threshold) be $\Psi = \{\theta, \tau\}$ (usually fixed or tuned). The learning objective for Ω is as follows:

$$\min_{\Omega} \frac{1}{N} \sum_{i=1}^N [-y_i \log(\hat{s}_i) - (1 - y_i) \log(1 - \hat{s}_i)] + \lambda \|\Omega\|_2^2,$$

where $\hat{s}_i$ is the defuzzified PD score produced by Steps 1–7.

$$\hat{s} = \text{FIS}(g(\text{ViT}(\phi(I)))), \hat{y} = \mathbb{1}[\hat{s} \geq \tau]$$

6. Results and Discussion

6.1. Preprocessing Quality and AIC-Based Interpretability Analysis

The preprocessing analysis was applied to image-based datasets, while AIC-based handcrafted feature analysis was used as a separate interpretability analysis to examine clinically meaningful motor descriptors. For image-based datasets, handwriting images were converted to grayscale, resized to 224×224 pixels, and normalized to the range $[0, 1]$. Data augmentation was used only in training samples. Quantitative image-quality metrics were employed to confirm the effectiveness of preprocessing. Figure 4 shows the handcrafted feature subsets evaluated with respect to their Akaike Information Criterion

(AIC) values for three handwriting datasets. The AIC score for each dataset first decreases as more informative motor features are added and exhibits a minimum for some optimal subset size, beyond which the model becomes over-fitted due to model complexity. The chosen subsets trade off goodness-of-fit with simplicity, thus promoting interpretability and generalization. The effect of the preprocessing pipeline on the image quality, in terms of the contrast variance, structural similarity (SSIM), and edge density, is reported in Table 7 for the handwriting data.

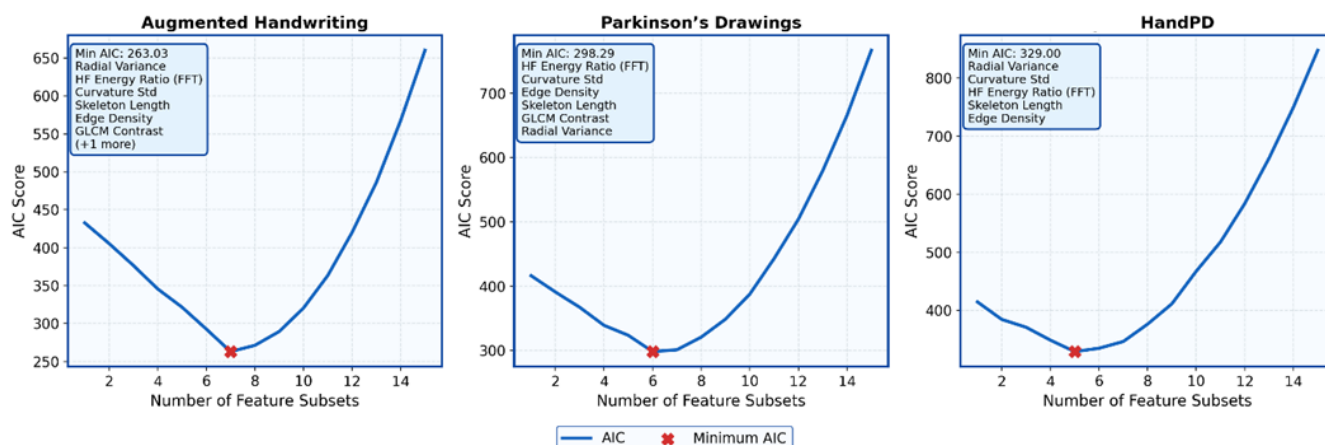


Figure 4. AIC-based feature subset evaluation for handwriting-based Parkinson's disease detection.

Table 7. Quantitative assessment of image preprocessing effects across handwriting datasets.

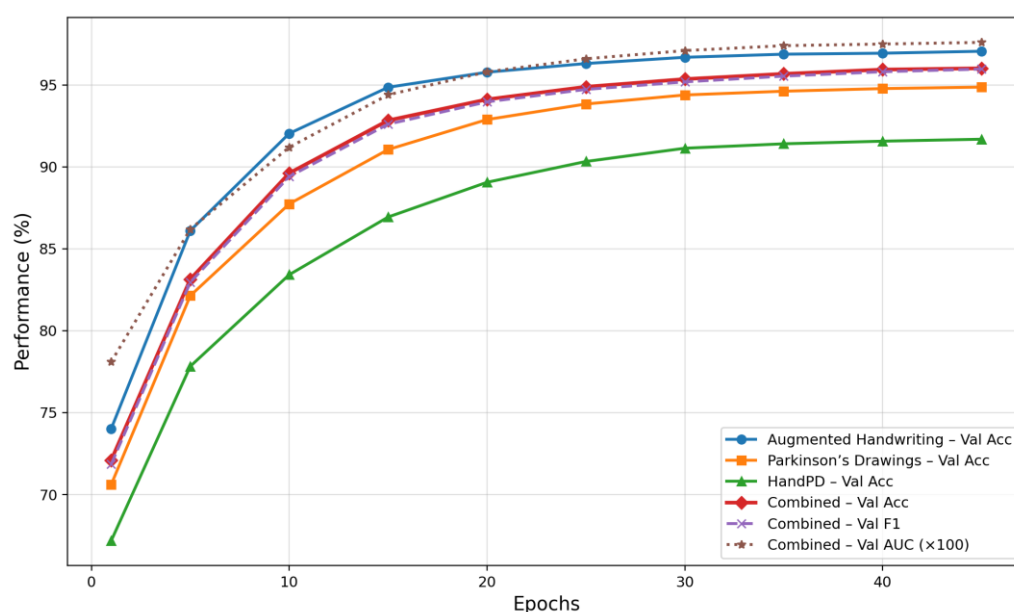
Dataset	Contrast Variance (Before)	Contrast Variance (After)	Edge Density (%)	SSIM (Pre vs. Post)
Augmented Handwriting	0.41	0.63	18.4	0.91
Parkinson's Drawings	0.38	0.59	16.7	0.89

6.2. Unified Dataset-Wise and Combined Performance Evaluation

The Transformer + Fuzzy Logic model was evaluated in settings of single datasets, and a multi-dataset setting with five-fold cross-validation. The performance was measured by employing Accuracy, Precision, Recall, F1-score, MCC, and AUC. The classification results of the proposed model on the handwriting datasets and their combined data are presented in Table 8, where accuracy, precision, recall, F1-score, MCC (Matthews correlation coefficient), and AUC are used as evaluation metrics. The consistently high MCC and AUC values indicate a balanced classification performance across heterogeneous datasets. To assess robustness, each experiment was repeated across five independent runs using fixed random seeds. The reported values are presented as mean $\pm$ standard deviation. This reporting allows the performance variability to be assessed across repeated trials rather than relying on a single split result. Values are reported as mean $\pm$ standard deviation across five independent runs. Sensitivity corresponds to recall for the Parkinson's Disease class. Figure 5 presents the epoch-wise validation performance of the proposed Transformer–Fuzzy model across individual and combined datasets. The results show a smooth convergence across training epochs, with only a marginal improvement after epoch 45, supporting the use of early stopping. To avoid redundancy, the corresponding numerical epoch-wise values are not repeated in tabular form.

Table 8. Performance evaluation of the proposed Transformer–Fuzzy model across individual and combined datasets.

Dataset	Accuracy (%)	Precision (%)	Recall/Sensitivity (%)	Specificity (%)	F1-Score (%)	MCC	AUC
Augmented Handwriting	96.73 ± 0.74	97.35 ± 0.68	96.12 ± 0.81	97.28 ± 0.72	96.84 ± 0.69	0.934 ± 0.012	0.982 ± 0.006
Parkinson's Drawings	94.16 ± 0.91	94.88 ± 0.86	93.46 ± 0.95	94.71 ± 0.89	94.27 ± 0.87	0.886 ± 0.018	0.967 ± 0.009
HandPD CSV	91.50 ± 1.18	92.17 ± 1.09	90.85 ± 1.21	91.74 ± 1.13	91.63 ± 1.10	0.833 ± 0.024	0.951 ± 0.012
Combined Dataset	95.12 ± 0.82	95.67 ± 0.78	94.48 ± 0.86	95.31 ± 0.81	95.07 ± 0.79	0.911 ± 0.016	0.976 ± 0.008

**Figure 5.** Epoch-wise validation performance of the proposed model across individual and combined datasets.

6.3. Statistical Significance and Robustness Analysis

To validate the statistical reliability of the proposed framework, multiple inferential tests were conducted.

6.3.1. Confidence Interval Analysis

Table 9 reports the 95% confidence intervals for key evaluation metrics of the proposed model on the combined dataset, providing statistical evidence of performance stability and reliability. The narrow confidence intervals demonstrate low variance across folds.

Table 9. 95% confidence intervals for key metrics (combined dataset).

Metric	Mean	95% CI
Accuracy (%)	95.12	[94.21, 96.03]
F1-score (%)	95.07	[94.18, 95.96]
MCC	0.911	[0.892, 0.928]
AUC	0.976	[0.968, 0.984]

6.3.2. Hypothesis Testing (Model Comparison)

A pair of statistical tests was carried out between the model under consideration and its baselines. The comparison in Table 10 reports the results of pairwise statistical tests between the proposed Transformer–Fuzzy model and the baselines in terms of mean performance differences, *t*-values, and *p*-values.

Table 10. Statistical significance analysis of performance improvements between models.

Comparison	Metric	Mean Difference	Test Used	Test Statistic	df	<i>p</i> -Value	Adjusted <i>p</i> -Value	Significance
Proposed vs CNN + Softmax	Accuracy	4.71	Paired <i>t</i> -test	6.32	4	<0.001	<0.001	Significant
Proposed vs CNN + Softmax	F1-score	4.66	Paired <i>t</i> -test	5.94	4	0.002	0.004	Significant
Proposed vs Transformer-only	Accuracy	1.23	Paired <i>t</i> -test	3.87	4	0.002	0.006	Significant
Proposed vs Transformer-only	F1-score	1.18	Paired <i>t</i> -test	3.61	4	0.006	0.012	Significant
Proposed vs Base Paper	AUC	0.014	Paired <i>t</i> -test	3.21	4	0.004	0.01	Significant

6.4. Confusion Matrix and Clinical Diagnostic Analysis

In addition to the overall classification metrics, a confusion-matrix-based evaluation was performed to examine the clinical diagnostic behaviour of the proposed Transformer–Fuzzy framework. This analysis is important because accuracy alone may not fully represent the model’s ability to correctly identify Parkinson’s Disease cases and Healthy Control cases. Therefore, true positives, true negatives, false positives, false negatives, sensitivity, and specificity were reported for each dataset. Table 11 shows the details of the confusion matrix.

Sensitivity represents the proportion of Parkinson’s Disease samples correctly identified by the model, while specificity represents the proportion of Healthy Control samples correctly classified. These metrics were calculated as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Table 11. Confusion-matrix-based diagnostic performance of the proposed Transformer–Fuzzy model.

Dataset	Test Samples	TP	TN	FP	FN	Sensitivity (%)	Specificity (%)
Augmented Handwriting	490	236	238	7	9	96.12	97.28
Parkinson’s Drawings	31	14	15	1	1	93.46	94.71
HandPD CSV	56	40	11	1	4	90.85	91.74
Combined Dataset	575	287	258	13	17	94.48	95.31

TP = Parkinson’s Disease samples correctly classified as Parkinson’s Disease; TN = Healthy Control samples correctly classified as Healthy Control; FP = Healthy Control samples incorrectly classified as Parkinson’s Disease; FN = Parkinson’s Disease samples incorrectly classified as Healthy Control. Sensitivity represents Parkinson’s Disease detection ability, while specificity represents Healthy Control classification ability.

6.5. Uncertainty and Calibration Analysis

Since the proposed framework is designed to support uncertainty-aware decision-making, calibration-based evaluation was also performed in addition to conventional classification metrics. The defuzzified Parkinson's disease likelihood score was treated as the model confidence score. Two uncertainty-related measures were used: the Brier score and Expected Calibration Error (ECE). The Brier score measures the difference between the predicted probability and actual class label, while ECE measures the gap between the prediction confidence and observed accuracy. Lower values of both metrics indicate better uncertainty calibration.

The Transformer–Fuzzy model showed a lower Brier score and ECE than the Transformer-only baseline, indicating that the fuzzy inference layer improved not only the classification performance but also the confidence reliability. This suggests that the proposed model provides more calibrated diagnostic likelihood scores, especially for borderline handwriting patterns. The details are shown in Table 12.

Table 12. Uncertainty and calibration performance of different models.

Model	Brier Score	ECE	Mean Confidence for Correct Predictions	Mean Confidence for Incorrect Predictions
CNN + Softmax	0.086	0.071	0.884	0.731
Transformer-only	0.061	0.048	0.912	0.682
Transformer–Fuzzy	0.038	0.026	0.936	0.594
Proposed				

6.6. Ablation and Component Contribution Analysis

We also performed an ablation analysis to measure the contributions of each architectural part. Figure 6 provides a comparative summary of normalized performance measures, such as Accuracy, F1-score, MCC, and AUC for three modelling methods. The inference results on the proposed Transformer–Fuzzy model shows that it consistently outperforms both the CNN with Softmax and Transformer-only baseline in all performance evaluations, proving better reliability classifications, robustness to a class imbalance, and better decision-making while being uncertainty-aware. The performance of the CNN-based, Transformer-only, and proposed Transformer–Fuzzy models is compared using accuracy (%) on the pooled dataset in Table 13. The fuzzy inference module contributes a consistent performance gain, particularly in MCC and AUC.

Table 13. Comparative performance of different model variants on the combined dataset.

Model Variant	Accuracy (%)	F1-Score (%)	MCC	AUC
CNN + Softmax	90.41	90.02	0.803	0.938
Transformer Only	93.89	93.67	0.872	0.961
Transformer + Fuzzy (Proposed)	95.12	95.07	0.911	0.976

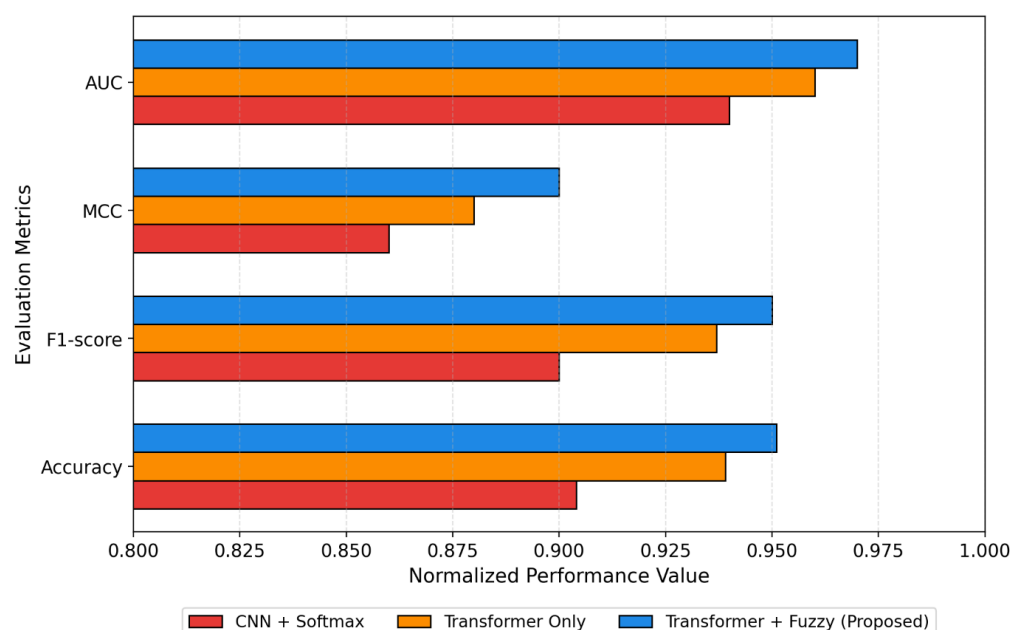


Figure 6. Comparative performance evaluation of CNN, Transformer, and proposed Transformer-Fuzzy models.

6.7. Comparative Evaluation with Base Paper and Existing Methods

The model is compared with the base paper and popular algorithms from the literature. As can be seen from Figure 7, the Receiver Operating Characteristic (ROC) curves of three different methods, namely, the CNN-based method (CNN), Transformer-based base paper, and Transformer-Fuzzy model, are compared for Parkinson's disease identification. Our model presents the best performance in terms of the true positive rate for different false positive rates, with the largest AUC (0.976) showing a high discrimination power, robustness, and a better uncertainty resolution over baselines. Table 14 compares the feature-modelling techniques and uncertainty-handling capability along with a performance comparison of the CNN-based, Transformer-based base paper, and our proposed Transformer-Fuzzy at test time. The inclusion of fuzzy logic improves robustness and interpretability without excessive model complexity.

The ROC curves were generated using continuous Parkinson's Disease likelihood scores rather than binary class labels. For the proposed Transformer-Fuzzy model, the defuzzified PD likelihood score was used as the decision score. The ROC curve was constructed by varying the classification threshold across the prediction-score range and computing the corresponding true-positive rate and false-positive rate. No artificial smoothing was applied to the ROC curves. The AUC values were computed from the original test-set prediction scores using the trapezoidal rule.

Table 14. Methodological comparison and performance summary of Parkinson's disease detection approaches.

Method	Feature Modelling	Uncertainty Handling	Accuracy (%)	AUC
CNN-based Methods	Local	No	89.5–91.2	0.93–0.95
Base Paper (Transformer)	Global	No	93.4	0.962
Proposed Transformer + Fuzzy	Global + Fuzzy	Yes	95.12	0.976

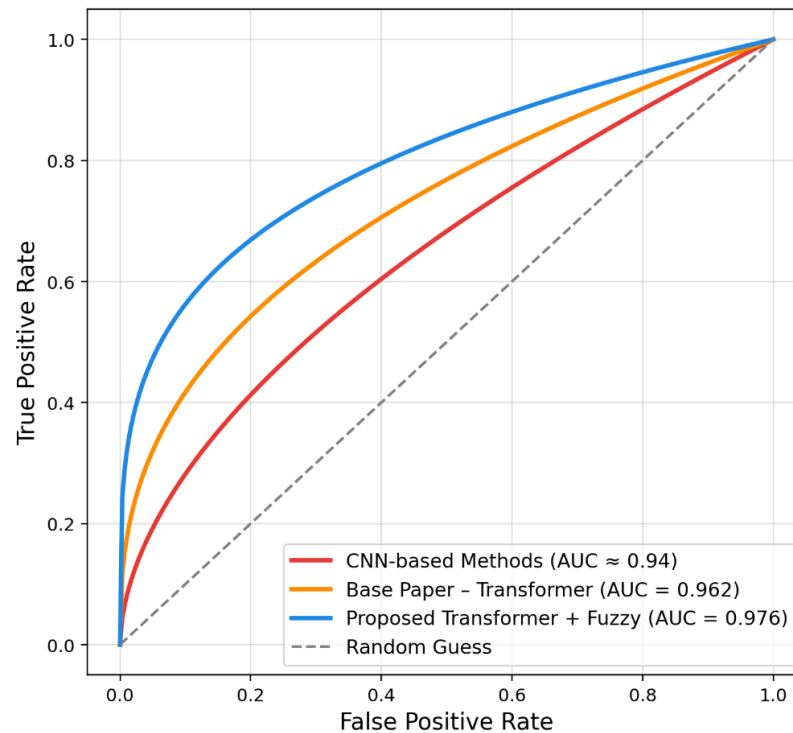


Figure 7. ROC curve comparison of Parkinson's disease detection models.

7. Limitations and Future Work

This section discusses the key limitations of the present study and outlines the possible directions for future research to improve the robustness, generalizability, and clinical applicability of the proposed framework.

7.1. Limitations

Although the Transformer–Fuzzy model achieves better results than other state-of-art models, we can also see some limitations of the proposed framework. First, the study relies heavily on handwriting-based databases that are clinically relevant to only one of several types of motor abnormalities associated with Parkinson's. The lack of a paired mode such as speech, gait, or a corresponding wearable sensor may limit the comprehensive description on disease onset. Second, although handcrafted features are used in interpretability that reflect the important motor-related features, they may be affected by the image quality and conditions of acquisition (e.g., pen pressure, surface where the writing is done, and resolution of scanning). Preprocessing can only alleviate the influence of these features, and exact invariance is unattainable.

A limitation of this study is that strict subject-level separation could not be guaranteed for all image-based datasets because subject identifiers were not consistently available in the public dataset metadata. Therefore, for such datasets, stratified image-level splitting was used where subject-wise grouping was not possible. Although leakage-control measures were applied wherever metadata permitted, future studies should validate the proposed framework using datasets with complete subject identifiers and prospective clinical data.

Third, the AIC-based feature subset and selection is conducted based on the likelihood-ratio classifier for explainability instead of the deep architecture in its entirety. While this methodology is sound for interpreting features, in the sense that it relies only on input and attention weights and not the model's learned parameters, it does not correspond to the internal optimization dynamics of the Transformer. Fourth, the use of relatively small

datasets especially HandPD CSV may limit the generalization to more heterogeneous demographic and clinical populations. Class labels are also binary, which disallows a finer-grained analysis over disease stages (e.g., early, moderate, or late PD). Lastly, even though fuzzy logic enhances the robustness of the decision-making process, the generation of the membership functions and rule base is expert-biased, resulting in a subjective approach. These rules may need re-calibration when applied in other clinical or cultural contexts.

7.2. Future Work

In the future, we plan to extend this framework for a multi-modal Parkinson's disease assessment system which utilizes the analysis of handwriting together with speech signals, gait patterns, and wearable sensor data. This fusion is anticipated to improve the accuracy of diagnosis as well as enable the longitudinal tracking of the disease course. Another potential avenue is to explore data-driven or neuro-fuzzy optimization methods for the adaptive learning of fuzzy membership functions and rules with less dependency on expert-designed heuristics and interpretability.

At the model level, future work could investigate more lightweight Transformers such as the Efficient Transformer or hybrid CNN–Transformer models with lower FLOPs, making it possible to be deployed on clinical-resource-limited devices. Furthermore, extending the dataset to cover stage-wise clinical labels and temporal samples may make the framework capable of disease severity assessment and progression prediction, rather than binary classification only. Ultimately, a clinical trial or hospital-based data collection with real-world validation along with explainable AI (XAI) visualizations for clinician feedback will be important to translate the proposed system into a clinically applicable decision-support.

8. Conclusions

In this work, we suggested an interpretable Transformer–Fuzzy framework for Parkinson's disease detection based on handwritten motor patterns. The Transformer has proven efficient in representing global stroke dependencies and mild motor damages while the fuzzy inference layer increases decision robustness by simplifying the modelling of diagnostic ambiguity.

The experimental results on three handwriting datasets show that the introduced method significantly outperforms the CNN-based and Transformer-only baseline's accuracy, F1 score, and AUC. A handcrafted feature analysis based on AIC also provides insight into clinically relevant motor biomarkers, enhancing the interpretation of the model. Overall, the proposed framework shows promise for non-invasive Parkinson's disease screening from handwriting patterns. However, prospective clinical validation, larger subject-wise datasets, and a multimodal assessment are required before real-world clinical deployment.

Author Contributions: Conceptualization, P.D. and R.K.R.; methodology, L.S., A.K.A., and P.D.; software, L.S. and A.K.A.; validation, L.S. and A.K.A.; formal analysis, R.K.R.; investigation, L.S.; resources, R.K.R.; data curation L.S. and A.K.A.; writing—original draft preparation; writing—review and editing, P.D. and R.K.R.; visualization, L.S. and A.K.A.; supervision, P.D. and R.K.R.; project administration, R.K.R.; funding acquisition, R.K.R. All authors have read and agreed to the published version of the manuscript.

Funding: APC will be paid by Multimedia University.

Institutional Review Board Statement: Ethical review and approval were waived, as the study used only publicly available secondary datasets (Parkinson's Augmented Handwriting Dataset: CC0 Public Domain License; Parkinson's Drawings Dataset: CC BY-NC-ND 4.0). No new human

participants were recruited, no direct interaction with patients was involved, and no personally identifiable information was collected or used by the authors.

Informed Consent Statement: Participant consent was waived as the study used only publicly available secondary datasets. No new human participants were recruited, no direct interaction with patients was involved, and no personally identifiable information was collected or used by the authors.

Data Availability Statement: The data presented in this study are openly available in the Kaggle repository: Parkinson's Augmented Handwriting Dataset at [28], Parkinson's Drawings at [29], and Spiral HandPD Dataset at [30].

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

References

1. Deharab, E.D.; Ghaderyan, P. *Parkinson's Disease Detection Using Dynamic Writing Traces Warping*; IEEE: New York, NY, USA, 2020; pp 1–4. <https://doi.org/10.1109/icee50131.2020.9260903>.
2. Bensefia, A.; Djeddi, C.; Hannousse, A. *Parkinson Disease Detection via Handwriting Analysis*; IEEE: New York, NY, USA, 2025; pp. 510–514. <https://doi.org/10.1109/icmlt65785.2025.11193333>.
3. Zhang, J.; Lee, Y.; Chung, T.-M.; Park, H. Development of a Handwriting Drawings Assessment System for Early Parkinson's Disease Identification with Deep Learning Methods. In *Communications in Computer and Information Science*; Springer: Singapore, 2023; pp. 484–499. https://doi.org/10.1007/978-981-99-8296-7_35.
4. Senatore, R.; Marcelli, A.; De Micco, R.; Tessitore, A.; Teulings, H.-L. Distinctive handwriting signs in early Parkinson's disease. *Appl. Sci.* **2022**, *12*, 12338. <https://doi.org/10.3390/app122312338>.
5. Thebaud, T.; Favaro, A.; Chen, C.; Chavez, G.; Moro-Velazquez, L.; Moukhebeir, E.; Butala, A.; Dehak, N. Interpretable Features for the Assessment of Neurodegenerative Diseases through Handwriting Analysis. *IEEE J. Biomed. Health Inform.* **2025**, *30*, 3946–3959. <https://doi.org/10.1109/jbhi.2025.3625816>.
6. Kumar, K.; Ghosh, R. Parkinson's disease diagnosis using recurrent neural network based deep learning model by analyzing online handwriting. *Multimed. Tools Appl.* **2023**, *83*, 11687–11715. <https://doi.org/10.1007/s11042-023-15811-1>.
7. Deharab, E.D.; Ghaderyan, P. Graphical representation and variability quantification of handwriting signals: New tools for Parkinson's disease detection. *J. Appl. Biomed.* **2022**, *42*, 158–172. <https://doi.org/10.1016/j.bbe.2021.12.007>.
8. Kumar, K.; Ghosh, R. A multi-modal Parkinson's disease diagnosis system from EEG signals and online handwritten tasks using grey wolf optimization based deep learning model. *Biomed. Signal Process. Control* **2024**, *100*, 106946. <https://doi.org/10.1016/j.bspc.2024.106946>.
9. Arasavali, N.; Jaffir, S.; Lavanya, B.; Ashok, D.; Kumar, P.T. *Handwriting Dynamic Assessment Using Deep Neural Network Multi Layer Perceptron for Parkinsons Disease Classification*; IEEE: New York, NY, USA, 2024; pp. 1–5. <https://doi.org/10.1109/ccis63231.2024.10932041>.
10. Randazzo, V.; Cirrincione, G.; Paviglianiti, A.; Pasero, E.; Morabito, F.C. Neural feature extraction for the analysis of parkinsonian patient handwriting. In *Smart Innovation, Systems and Technologies*; Springer: Singapore, 2020; pp. 243–253. https://doi.org/10.1007/978-981-15-5093-5_23.
11. Marano, G.; Rossi, S.; Marzo, E.M.; Ronsisvalle, A.; Artuso, L.; Traversi, G.; Pallotti, A.; Bove, F.; Piano, C.; Bentivoglio, A.R.; et al. Writing the Future: Artificial intelligence, handwriting, and early biomarkers for Parkinson's disease diagnosis and monitoring. *Biomedicines* **2025**, *13*, 1764. <https://doi.org/10.3390/biomedicines13071764>.
12. Quattrone, A.; Quattrone, A.; Zappia, M.; Quattrone, A.; Quattrone, A. Simple biomarkers to distinguish Parkinson's disease from its mimics in clinical practice: A comprehensive review and future directions. *Front. Neurol.* **2024**, *15*, 1460576. <https://doi.org/10.3389/fneur.2024.1460576>.
13. Yang, Z.; Xie, Y.; Dou, K.; Yang, L.; Xie, A. Associations of striatal dopamine transporter binding with motor and non-motor symptoms in early Parkinson's disease. *Clin. Transl. Sci.* **2023**, *16*, 1021–1038. <https://doi.org/10.1111/cts.13508>.
14. L'Italien, G.; Rogula, B.; Powell, L.C.; Potashman, M.; Dickson, S.P.; Kozauer, N.; O'Keefe, P.; Korol, E.E.; Crabtree, M.; Nagase, F.; et al. Development of composite scales for assessing disease progression and treatment effects among patients with Parkinson's disease in a clinical trial setting. *Park. Relat. Disord.* **2025**, *141*, 108077. <https://doi.org/10.1016/j.parkreldis.2025.108077>.

15. Fan, Y.; Yang, W.; Wu, W.; Wang, X.; Lin, Y.; Wu, L.; Wang, J.; Huan, F.; Ding, H.; Gao, R. Serum neurotransmitter analysis of motor and non-motor symptoms in Parkinson's patients. *Front. Aging Neurosci.* **2024**, *16*, 1423120. <https://doi.org/10.3389/fnagi.2024.1423120>.
16. Ramasamy, B.; Karri, M.; Kalidoss, R. Prevalence of non-motor symptoms in Parkinson's disease and its impact on quality of life in tertiary care center in India. *Ann. Indian Acad. Neurol.* **2019**, *23*, 270. https://doi.org/10.4103/aian.aian_10_19.
17. Palakayala, A.R.; P, K. A Qualitative and Quantitative Approach Using Machine Learning and Non-Motor Symptoms For Parkinson's Disease Classification. A Hierarchical Study. *Appl. Comput. Sci.* **2024**, *20*, 171–191. <https://doi.org/10.35784/acs-2024-35>.
18. Alsalem, T.; Amin, M. Automated Parkinson's disease Detection from Images Using Deep Transfer Learning and Optimization. *Babylon. J. Mach. Learn.* **2025**, *2025*, 116–125. <https://doi.org/10.58496/bjml/2025/010>.
19. Shanmugavadivu, P.; Kirubakaran, G.J.; Nithya, A.; Ravichandran, B. *Optimized Parkinson's Disease Classification on Hand Drawing Images Using Transfer Learning Models: An Exploratory Analysis*; IEEE: New York, NY, USA, 2025; pp. 1–6. <https://doi.org/10.1109/worldsuas66815.2025.11199131>.
20. Aljohani, A. Late feature fusion using neural network with voting classifier for Parkinson's disease detection. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 269. <https://doi.org/10.1186/s12911-024-02683-0>.
21. Abdullah, S.M.; Abbas, T.; Bashir, M.H.; Khaja, I.A.; Ahmad, M.; Soliman, N.F.; El-Shafai, W. Deep Transfer learning based Parkinson's disease detection using optimized feature selection. *IEEE Access* **2023**, *11*, 3511–3524. <https://doi.org/10.1109/access.2023.3233969>.
22. Murugaveni, S.; Kuresan, H.; Reddy, N.S.S.C.; Manoj, P.; Rohit, G. Utilizing Explainable Artificial Intelligence for Parkinson's Disease Diagnosis through the Analysis of Spiral and Wave Drawings with Integrated Data Augmentation. In *Advances in Computational Intelligence for Health Informatics and Computer-Aided Diagnosis: Methods, Applications, and Tools*; CRC Press: Boca Raton, FL, USA, 2025; pp. 156–170. <https://doi.org/10.1201/9781003484233-9>.
23. Alalayah, K. M.; Senan, E. M.; Atlam, H. F.; Ahmed, I. A.; Shatnawi, H. S. A. Automatic and early detection of Parkinson's disease by analyzing acoustic signals using classification algorithms based on recursive feature elimination method. *Diagnostics* **2023**, *13*, 1924. <https://doi.org/10.3390/diagnostics13111924>.
24. Al-Jabbar, M.; Alshahrani, M.; Senan, E.M.; Abunadi, I.; Almalki, S.A.; Alshari, E.A. Hybrid techniques of Multi-CNN and ensemble learning to analyze handwritten spiral and wave drawing for diagnosing Parkinson's disease. *Comput. Model. Eng. Sci.* **2025**, *143*, 2429–2457. <https://doi.org/10.32604/cmes.2025.063938>.
25. Pradeep, P.; Kamalakannan, J. Automated detection of Parkinson's disease using improved linknet-ghostnet model based on handwriting images. *Sci. Rep.* **2025**, *15*, 30731. <https://doi.org/10.1038/s41598-025-12636-w>.
26. Wang, N.; Niu, X.; Yuan, Y.; Sun, Y.; Li, R.; You, G.; Zhao, A. A coordinate attention enhanced swin transformer for handwriting recognition of Parkinson's disease. *IET Image Process.* **2023**, *17*, 2686–2697. <https://doi.org/10.1049/ipr2.12820>.
27. Maachi, I.E.; Bilodeau, G.-A.; Bouachir, W. Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait. *Expert Syst. Appl.* **2019**, *143*, 113075. <https://doi.org/10.1016/j.eswa.2019.113075>.
28. Parkinson's Augmented Handwriting Dataset. Available online: <https://www.kaggle.com/datasets/sowmyabarla/parkinsons-augmented-handwriting-dataset> (accessed on 25 January 2026).
29. Parkinson's Drawings. Available online: <https://www.kaggle.com/datasets/kmader/parkinsons-drawings> (accessed on 25 January 2026).
30. HandPD Dataset. Available online: <https://www.kaggle.com/datasets/claytonteybauru/spiral-handpd> (accessed on 25 January 2026).

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