Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 11, No. 03, September 2025.

HWR-PDNET: A TRANSFER LEARNING CNN FOR PARKINSON'S DETECTION FROM HANDWRITING IMAGES

Mathu T.[1], Ronal Roy[2], Jenefa Archpaul[2] and Ebenezer V.[2]

(Received: 6-Jun.-2025, Revised: 13-Aug.-2025, Accepted: 22-Aug.-2025)

ABSTRACT

Parkinson's Disease (PD) is a progressive, chronic neurological disorder that is distinguished by abnormalities in the motor system. The condition can be detected in the early stage by the irregular handwriting of the individual. Early diagnosis is critical to enable timely therapeutic intervention and slow disease advancement. However, traditional diagnostic approaches largely depend on subjective clinical assessments, which lack scalability and exhibit reduced sensitivity in the prodromal phase. The present study proposes a well-established deep-learning architecture using transfer learning with MobileNetV2, which can be used for early diagnosis of Parkinson's Disease through handwriting images. The dataset includes 816 samples from 120 people. It is augmented through grayscale and HSL to add more variety to feature samples of the model. A two-stage training regimen—initial base freezing followed by fine-tuning with a reduced learning rate—was employed to optimize convergence and generalization. The approach presented in this study scored 92% on accuracy with an F1-score of 0.88 and a precision of 0.81, outperforming those of conventional baselines in regard to sensitivity and robustness. The resulting framework is lightweight, non-invasive, and well-suited for real-time screening applications, offering significant potential for clinical decision support and remote telehealth deployments.

KEYWORDS

Parkinson's disease, Handwriting analysis, Early diagnosis, Transfer learning, Deep learning in healthcare, Convolutional neural networks (CNNs).

1. INTRODUCTION

Parkinson's disease is characterized by the loss of dopaminergic neurons in the Substantia Nigra (SN) region of the brain [1]-[2]. Tremors, slow movement, muscle stiffness, and balance difficulties are the clinical features that impact the affected person's daily-living capacity and quality of life. Due to the growing global incidence of these diseases, especially in aging populations, early and accurate diagnosis is a crucial goal of neurological care [4]-[5]. Nonetheless, there is a significant clinical issue with the early detection of Parkinson's disease. Currently, diagnosing a patient often involves identifying neurological symptoms and examining motor symptoms, but it is a subjective measure to start with and will result in late-stage diagnosis. Also, in resource-limited settings or pediatric and early-onset cases, these methods might lead to the delay or misdiagnosis of the medical intervention due to symptomatological similarities with other conditions or the atypical nature of the PD course [6]-[8]. Traditional methods have aimed to assist in the diagnosis of PD through the application of biomedical signals, including speech recordings, handwriting dynamics, and neuroimaging data. Although these methods are promising, they rely on features designed by hand, domain expertise, and hand-crafted pre-processing. Their performance also tends to drop off in real-world deployments and cross-population settings, limiting scalability and clinical utility [9]-[10]. To resolve this issue, the study introduces an efficient deep-learning framework that uses convolutional neural network (CNN) transfer learning to detect Parkinson's disease early in handwriting. Writing, a fine motor skill, may be affected by micrographia and other altered stroke patterns at the onset of Parkinson's disease because of micrographia. With a pre-trained MobileNetV2 model and data augmentations, our model improves feature-extraction capability with fewer data and computations. This research's principal contributions are outlined as follows:

- We present a deep-learning (DL) framework based on transfer learning using MobileNetV2 for early PD detection from handwriting images.

1. Mathu T. is with Nehru Institute of Technology, Coimbatore, India. Email: mathumetilda.t@gmail.com
2. Ronal Roy, Jenefa Archpaul and Ebenezer V. (Corresponding author) are with Karunya Institute of Technology and sciences, Coimbatore, India. Emails: {ronalroy, jenefaa, ebenezerv}@karunya.edu

- A robust data-enhancement pipeline using grayscale and HSL transformations is proposed to increase the diversity of the dataset.
- The suggested model delivers a classification accuracy of 92%, a strong performance for non-invasive PD screening in large populations.

The remaining sections of this study are organized as follows: Section 2 explores the recent advances in AI-enabled PD diagnosis. The proposed method explained in Section 3 consists of dataset pre-processing, CNN architecture, and training strategies. In Section 4, the results of the experiment and their comparison will be highlighted. Ultimately, Section 5 concludes the study and indicates future research directions.

2. RELATED WORK

Recent advancements in DL have significantly improved the diagnostic capabilities for Parkinson's Disease (PD) across various modalities. Alissa et al. (2021) [1] developed a CNN-based model utilizing figure-copying tasks, such as cube and pentagon drawings, achieving high accuracy by analyzing geometric distortions linked to PD. Similarly, Hireš et al. (2021) [2] introduced an ensemble of CNN models for detecting PD from voice recordings by leveraging acoustic features, such as pitch and jitter, yielding 90% accuracy. Chen et al. (2024) [3] proposed a CNN–Transformer hybrid network for segmenting PD-related nuclei from medical images, enhancing segmentation performance through long-range dependency modeling. Aggarwal et al. [4] suggested a one-dimensional convolutional neural-network framework with data augmentation to differentiate Parkinson's disease from SWEDD scans, yielding favorable classification outcomes. Wang et al. (2024) [5] compared 1D, 2D, and 3D CNNs for classifying digitized drawing tests, showing that dimensionality affects diagnostic performance in handwriting-based PD detection.

Focusing on motor-skill degradation, Allebawi et al. (2024) [6] implemented a handwriting-based PD detection system using a Beta-Elliptical model and fuzzy perceptual detectors, emphasizing dynamic spatiotemporal signatures in writing. For gait-related symptoms, Sigcha et al. (2024) [7] evaluated DL algorithms across datasets for freezing of gait (FoG) detection, highlighting the importance of standardization for clinical use. In the auditory domain, Celik and Başaran (2023) [8] presented a CNN–Random Forest hybrid model for PD detection using speech signals, showcasing robustness in feature modeling. Extending this, Madusanka and Lee (2024) [9] utilized transformer-based models on spectrograms of speech data, achieving 90.8% accuracy by identifying vocal biomarkers indicative of PD.

EEG-based approaches have also gained attention. Khalid and Ehsan (2024) [10] used gated recurrent units to classify EEG sub-bands, capturing temporal dependencies in brain activity related to PD and achieving notable accuracy. From an algorithmic perspective, Li et al. (2021) [11] provided an extensive survey on CNNs, covering applications across biomedical domains. Image pre-processing is essential in medical imaging; Qi et al. (2021) [12] provided a comprehensive overview of enhancement techniques, while van Dyk and Meng (2001) [13] discussed the statistical underpinnings of data augmentation to improve generalization.

In feature representation, Ping (2013) [14] reviewed classical image feature extraction methods, laying the groundwork for more complex deep-learning features. For lightweight CNN design, Dong et al. (2020) [15] introduced MobileNetV2, which balances efficiency and performance—making it suitable for PD detetion on limited data. For activation functions, He et al. (2018) [16] explored the theoretical foundations of ReLU in deep neural networks. Optimization strategies were improved by Zhang (2018) [17], who proposed an enhanced Adam optimizer for faster convergence. Transfer-learning techniques were thoroughly reviewed by Zhuang et al. (2020) [18], establishing their utility for domain adaptation, especially in healthcare. Radenović et al. (2016) [19] demonstrated unsupervised fine-tuning of CNNs using hard examples, supporting robust image retrieval and classification. Finally, Corley et al. (2015) [20] explored deep learning for software feature location, indirectly informing architecture search techniques relevant to model customization in PD-detection frameworks. Jiang et al. (2025) [21] proposed a novel network architecture specifically tailored for Parkinson's handwriting-image recognition, demonstrating enhanced structural modeling of handwriting patterns using domain-specific features. Extending this direction, Lu et al. (2025) [22] introduced a dynamic

handwriting feature-extraction approach that integrates temporal and spatial cues, showing significant improvement in Parkinson's disease-detection accuracy through dynamic pen-motion analytics. Kansizoglou et al. (2025) [23] contributed a hierarchical deep-learning framework that incorporates drawing-aware context to refine model performance, emphasizing the importance of spatial abstraction in analyzing handwriting traits linked to Parkinsonian symptoms. Miah et al. (2025) [24] conducted a comprehensive review encompassing various data modalities, including handwriting, voice, and motion signals, and highlighted structural and algorithmic considerations for future research on Parkinson's disease-detection systems. While most studies focus on Parkinson's-specific datasets, Javeed et al. (2025) [25] broadened the application domain by applying machine-learning techniques to classify handwriting samples for mental-health conditions, such as schizophrenia and bipolar disorder, underscoring the potential of handwriting as a universal biomarker for neurological and psychiatric evaluations. Al-Shannaq and Elrefaei [26] proposed a domain-specific transfer-learning method for age estimation. While existing methods demonstrate notable performance in Parkinson's Disease detection, many face limitations, such as restricted generalization on small datasets, insufficient stage-wise analysis, and limited use of domain-specific augmentations. These gaps motivate the proposed HWR-PDNet framework, which is designed to enhance robustness, improve early-stage detection, and address the shortcomings identified in prior approaches.

3. SYSTEM METHODOLOGY

This section elaborates on the suggested DL approach for the automated recognition of PD using handwriting image analysis. This method utilizes the representational capabilities of CNNs enhanced by transfer learning, enabling robust classification even with a limited dataset. The system comprises multiple stages, including image pre-processing, feature extraction, classification, and evaluation, as shown in Figure 1.

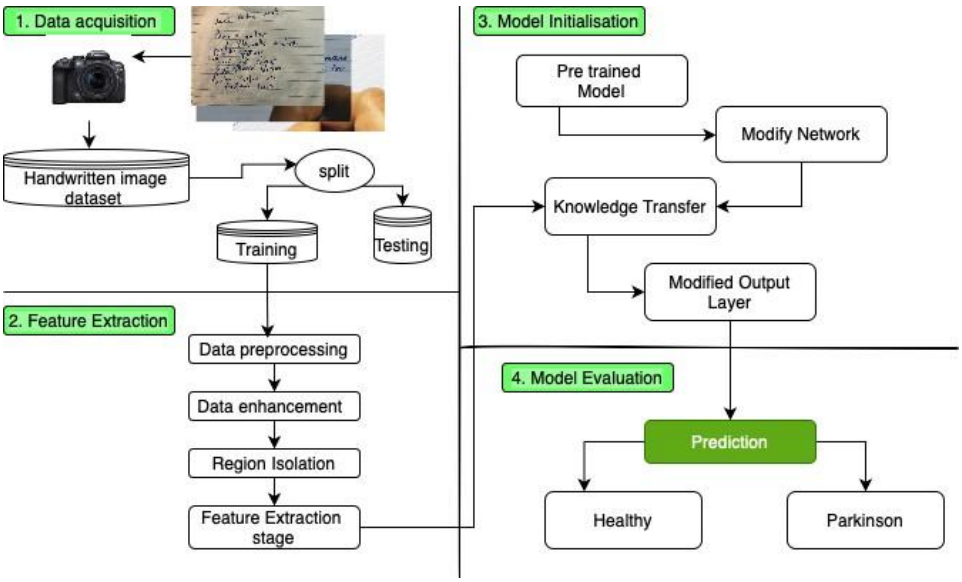


Figure 1. Workflow of the proposed CNN + transfer-learning system for PD recognition.

3.1 Data Acquisition

The handwriting-image dataset was compiled from both PD patients and healthy control participants. All subjects performed a standardized wave-drawing (saw-tooth) task using an identical pen-tablet device under controlled acquisition conditions. The captured images were then systematically split into training and testing sub-sets to facilitate model development and evaluation.

3.1.1 Problem Formulation

Let $D = \{(I_i, y_i)\}_{i=1}^{n}$ represent the dataset, where each handwriting image $I_i \in R^{H \times W \times C}$ corresponds to height H, width W, and C color channels (typically RGB, so C = 3). The label y ∈ {0, 1} denotes the ground-truth class, where 0 indicates a healthy individual and 1 corresponds to a patient diagnosed with PD. The objective is to learn a mapping function:

$$f_\theta : R^{H \times W \times C} \rightarrow \{0,1\} \quad (1)$$

where f_θ is a deep neural network parameterized by θ, which accurately classifies input images into one of the two target classes.

3.2 Feature Extraction

This stage prepares handwriting images for deep-learning input through pre-processing, augmentation, region isolation, and feature derivation. Initially, all images are resized to 256×256 pixels and normalized. Data-augmentation techniques—horizontal and vertical flips, ±20° rotations, zooming, and contrast adjustments—are applied to improve generalization. Grayscale conversion and HSL transformation emphasize stroke patterns and pen pressure variations. Region isolation reduces noise by focusing only on handwriting strokes. Finally, a pre-trained MobileNetV2 backbone is used to derive discriminative latent features.

3.3 Image Pre-processing and Augmentation

The collected handwriting images exhibit variability in image size and background noise. All images are resized to the 256×256 pixels. Data augmentation applies to a dataset of a restricted size and improves the model's generalization. This includes random horizontal and vertical flipping, rotations within ±20°, zooming, and contrast adjustments. Additionally, grayscale conversion and BGR to HSL transformation are incorporated to emphasize fine motor patterns and variations in pen pressure and stroke directionality — features often indicative of PD onset.

3.4 Feature Derivation Using Transfer Learning

In this work, we utilize a pre-trained lightweight deep network, symbolized as Ψ_{base}, originally optimized on the ImageNet benchmark, to perform feature derivation. For a given input handwriting image denoted by $X_n \in R^{H \times W \times C}$, the model outputs an intermediate feature embedding:

$$f_n = \Psi_{base}(X_n), \; f_n \in R^M \quad (2)$$

where f_n represents the extracted descriptor for sample *n*, and *M* is the latent vector dimensionality. This embedding captures both structural and abstract traits within the handwriting image that may be linked to Parkinsonian motor abnormalities.

3.5 Model Initialization

The MobileNetV2 backbone is adapted for binary classification by replacing its output layer with a task- specific classification head. Transfer learning is performed in two phases: first, freezing the backbone and training only the classification head at a learning rate of 10^{-4}; second, unfreezing the entire network and fine-tuning at a reduced learning rate of 10^{-5} to adapt the pre-trained features to the handwriting domain.

3.6 Decision Mapping Layer

The derived vector f_n is forwarded into a dense projection layer, followed by a softmax classifier to predict the output probabilities:

$$p_n = \text{softmax}(W_c \cdot f_n + b_c), \; p_n \in R^2 \quad (3)$$

Here, $W_c \in R^{2 \times M}$ and $b_c \in R^2$ denote the classification weights and bias terms. The predicted vector p_n reflects the confidence distribution across the binary output space, identifying whether the input sample is from a healthy subject or a PD patient.

3.7 Optimization Objective and Parameter Update

The network optimizes the sparse categorical cross-entropy loss between the actual labels y_n and predicted outputs p_n.

$$J_{CE} = -\sum_{j=1}^{2} y_{n,j} \log(p_{n,j}) \quad (4)$$

where $y_{n,j}$ and $p_{n,j}$ indicate the true label and predicted score for class j of the nth instance. To update model parameters ω, we employ the Adam optimizer with momentum-based adaptive learning. The parameter-update rule is defined as:

$$\boldsymbol{\omega}^{(u+1)} = \boldsymbol{\omega}^{(u)} - \lambda \cdot \nabla_{\boldsymbol{\omega}} \mathbf{J}_{CE} \tag{5}$$

where λ is the learning rate, and u indicates the current update step. Two separate learning rates are used: $\lambda = 10^{-4}$ for initial training (frozen backbone), and $\lambda = 10^{-5}$ for fine-tuning (unfrozen backbone).

3.8 Two-stage Training Framework

The training protocol consists of a dual-phase learning routine. In the preliminary phase, the base encoder Ψ_{base} is kept frozen to retain the pre-learned general features, while only the classifier head is trained on the PD-specific dataset. In the subsequent fine-tuning phase, the entire model including the feature extractor is unfrozen and optimized using a reduced learning rate. This two-stage strategy ensures efficient convergence and avoids overfitting, especially when working with limited domain-specific samples.

3.9 Non-linear Activation Dynamics

We incorporate non-linearity into the model by employing Rectified Linear Units (ReLUs) in hidden layers. This allows us to improve the learning capacity of the model. Given an input scalar s that is contained inside the set of real numbers, the ReLU activation is expressed as follows:

$$\mathrm{ReLU}(s) = \max(0, s) \tag{6}$$

This function suppresses negative activations and introduces sparsity, thereby improving gradient flow and learning stability. The transformed hidden output g is computed as:

$$\mathbf{g} = \mathrm{ReLU}(\mathbf{W}_h \cdot \mathbf{f}_n + \mathbf{b}_h) \tag{7}$$

where $\mathbf{W}_h$ and $\mathbf{b}_h$ are the parameters of the hidden fully connected layer.

3.10 Model Evaluation

Once trained, the model produces prediction probabilities for both PD and healthy classes. A confidence- based decision threshold τ is applied to balance sensitivity and specificity based on clinical-screening requirements. The model's performance is evaluated using accuracy, precision, recall, F1-score, and ROC AUC metrics.

3.11 Dropout-based Regularization Mechanism

To counteract overfitting due to the small sample size, a dropout mechanism is applied post-feature extraction. Let the dropout probability be denoted by ρ = 0.2, then the stochastic regularized output is computed as:

$$\tilde{\mathbf{g}} = \mathbf{g} \odot \boldsymbol{\delta}, \ \delta_i \sim \mathrm{Bernoulli}(1 - \rho) \tag{8}$$

Here, δ is a binary dropout mask applied element-wise using the Hadamard product ⊙. This introduces controlled noise during training, which improves model robustness by preventing reliance on specific neuron activations and enhancing generalization to unseen handwriting patterns.

3.12 Model Confidence and Decision Thresholding

The softmax output $\hat{y} = [\hat{y}_0, \hat{y}_1]$ represents the class probabilities for the two classes. The default decision rule assigns the class with the highest probability:

$$\hat{y}_{pred} = \arg\max \hat{y}_k, \ k \in \{0,1\} \tag{9}$$

However, to account for medical-risk tolerance, a confidence-based threshold $\tau \in [0,1]$ is introduced, such that:

$$\hat{y}_{pred} = \begin{cases} 1, & if\ \hat{y}_1 \geq \tau \\ 0, & otherwise \end{cases} \tag{10}$$

This allows tuning the sensitivity-specificity trade-off according to application needs, such as favoring early detection (high recall) over absolute precision in clinical-screening scenarios.

Algorithm 1 Proposed Parkinson's Disease Detection Pipeline

Require: Dataset D ={(I_i , y_i)}, pre-trained CNN $\varphi_{\text{MobileNet}}$, learning rates η_1, η_2, threshold τ

Ensure: Predicted labels $\hat{y}_i \in \{0,1\}$

1: **Preprocessing:** Resize I_i to 256×256, apply grayscale & HSL conversion, and augment with flip, rotation, zoom, contrast.

2: **Feature Extraction:** Compute $\mathbf{z}_i = \varphi_{\text{MobileNet}}(I_i)$

3: **Dropout Regularization:** $\tilde{\mathbf{z}}_i = \mathbf{z}_i \odot \mathbf{r}$, where $r_j \sim \text{Bernoulli}(1-p)$

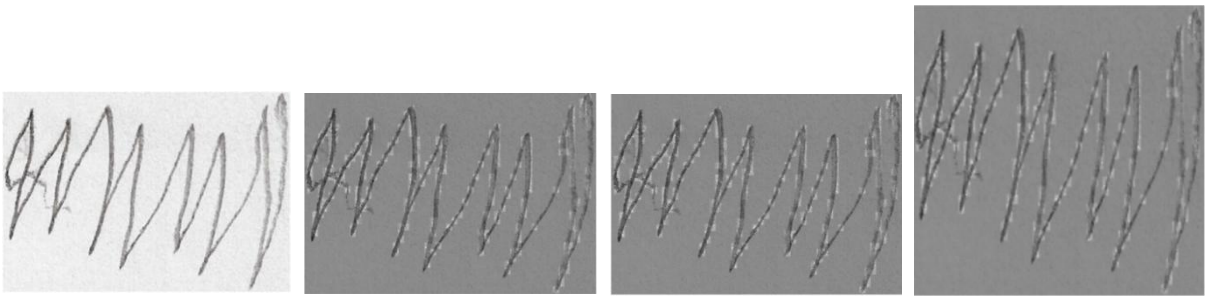
4: **Classification:** $\hat{\mathbf{y}}_i = \text{softmax}(W \cdot \tilde{\mathbf{z}}_i + \mathbf{b})$

5: **Loss:** $\mathcal{L}_{CE} = -\sum_{k=1}^{2} \boldsymbol{y}_{i,k} \boldsymbol{log}(\hat{\boldsymbol{y}}_{i,k})$

6: **Training:** Optimize θ using Adam with η_1 (frozen base); fine-tune with η_2 (unfrozen base)

7: **Prediction:**

$$\hat{y}_{pred} = \begin{cases} 1, & if\ \hat{y}_{i,1} \geq \tau \\ 0, & otherwise \end{cases}$$



(a) Initial image, (b) Grayscale enhanced image, (c) Unscaled image (1010×610), (d) Scaled image (256×256)

Figure 2. Progression of handwriting-image transformations: (a) Initial image, (b) grayscale enhancement, (c) original unscaled image and (d) resized image for CNN input.

4. RESULTS

4.1 Dataset Summary

The dataset used in this study was specifically curated to capture fine motor-skill anomalies typically observed in patients with Parkinson's Disease (PD), along with representative samples from neurologically healthy controls, as summarized in Table 1. A total of 120 subjects participated, comprising 60 clinically diagnosed PD patients and 60 healthy controls. The cohort included an equal gender distribution (60 males and 60 females) to ensure demographic balance, and the participants' ages ranged from 45 to 80 years, representing the most common age span for PD onset. The PD group was stratified according to the Hoehn and Yahr scale, a widely accepted clinical metric for disease severity: 20 patients in Stage 1 (early PD), 30 in Stage 2 (mild), 25 in Stage 3 (moderate), 20 in Stage 4 (severe), and 25 in Stage 5 (advanced). The healthy controls were screened to confirm the absence of neurological or movement disorders and were matched to the PD group by age and demographic background to minimize potential bias. All subjects performed a standardized wave-drawing (saw-tooth) task using the same pen-tablet device under uniform acquisition conditions, ensuring comparability of handwriting features. From these drawings, two primary kinematic attributes—pen pressure and drawing speed—were extracted, as they are clinically validated indicators of motor dysfunctions, such as micrographia, tremor, and bradykinesia.

The raw handwriting images were captured at an original image size of 1010×610 pixels and subsequently resized to 256×256 pixels to meet the input-dimensionality requirements of the MobileNetV2 architecture. The dataset was divided into training (80%), validation (10%), and testing (10%) sub-sets, maintaining proportional representation of PD stages and healthy controls in each split. To further increase intra- class diversity and improve generalization, data-augmentation techniques—including grayscale and HSL conversion, geometric transformations, and contrast enhancement—were applied. This process expanded the dataset to 816 images, enabling robust learning despite the relatively limited original sample size.

Table 1. Dataset description for Parkinson's disease handwriting study.

Attribute	Details
Total Subjects	120 (60 PD patients, 60 Healthy Controls)
Age Range	45–80 years
Primary Features	Pen-pressure, Drawing Speed
Image Size (Original)	1010 × 610 pixels
Image Size (Resized)	256 × 256 pixels
PD Stage Classification	Hoehn and Yahr Scale (Stage 1 to Stage 5)
Stage 1 (Early PD)	20 Patients
Stage 2 (Mild PD)	30 Patients
Stage 3 (Moderate PD)	25 Patients
Stage 4 (Severe PD)	20 Patients
Stage 5 (Advanced PD)	25 Patients
Healthy Controls	60 Subjects
Data Split	80% Train, 10% Validation, 10% Test
Final Dataset Size (after augmentation)	816 Images

4.1 Model Configuration and Hyper-parameter Settings

The suggested framework utilizes MobileNetV2 as a feature extractor, because it requires fewer resources to run and performs well in environments with limited power, as shown in Table 2. This model used pre-trained weights for ImageNet, allowing effective transfer learning to use its model for handwriting classification of people with Parkinson's disease. All writing samples were resized to 256×256×3 to conform with the input structure requirements of the model. A data augmentation pipeline was utilized to improve generalization and reduce overfitting. This step involved flipping images horizontally and vertically at random, rotating images up to 20°, zooming, and changing contrast. Each of these transformations was done with a probability of 0.2, enabling variability similar to real-world handwriting. After the convolutions, a Global Average Pooling (GAP) layer is utilized to lower the feature's dimension while obtaining a reduced characteristic map and compressing spatial information by bridging spatial features to obtain the most discriminative features. Then, a dropout layer with 20% drop probability was added before output dense layers to prevent neuron co-adaptation. The classification portion was made up of a fully connected layer composed of 64 units, each activated by the ReLU function, followed by a soft- max output to predict the probabilities of Parkinson's and Healthy. We utilized the Sparse Categorical Cross-Entropy objective function, appropriate for multi-class classification problems, including degenerate binary cases.

Table 2. Hyper-parameters and training configuration.

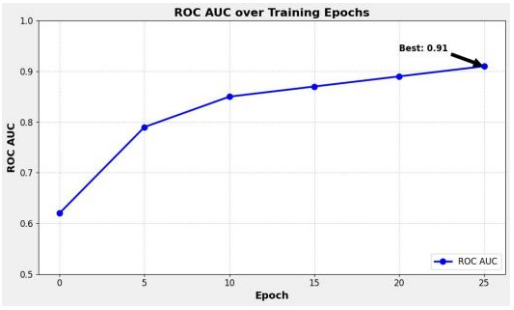
Hyper-parameter	Value
Base Model	MobileNetV2 (Pre-trained on ImageNet)
Input Image Size	256 × 256 × 3
Data Augmentation	Flip, Rotation (0.2), Zoom (0.2), Contrast (0.2)
Pooling Layer	Global Average Pooling
Dropout Rate	0.2
Dense Layer	64 Units, ReLU Activation
Output Layer	Softmax (2 Classes: Healthy / PD)
Loss Function	Sparse Categorical Cross Entropy
Optimizer (Initial Phase)	Adam (LR = 1e-4)
Optimizer (Fine-tuning Phase)	Adam (LR = 1e-5)
Batch Size	32
Total Epochs	25 (15 Base + 10 Finetuning)

Due to its adaptability to the gradients and speed of convergence, training was performed *via* the Adam optimization. To begin with training the classification head, the learning rate was set to 1×10^{-4}

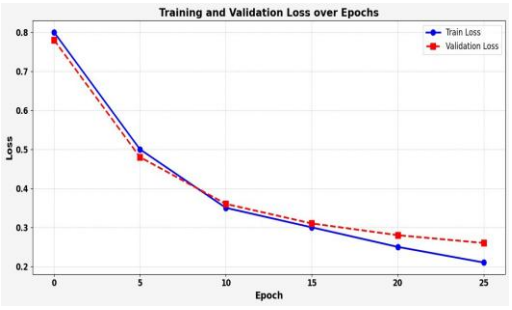
with the feature extractor frozen. In the fine-tuning step, all layers were unfrozen and re-optimized using a lower learning rate (1e−5). The model underwent training with a batch size of 32. In total, we ran 25 epochs, where we used 15 epochs for training and 10 for fine-tuning. The transformation of raw handwriting samples is illustrated in Figure 2. Certain image pre-processing operations are done to raw handwriting samples to edit and resize them for training.

Table 3. Metrics for training and validation by epoch using ROC AUC.

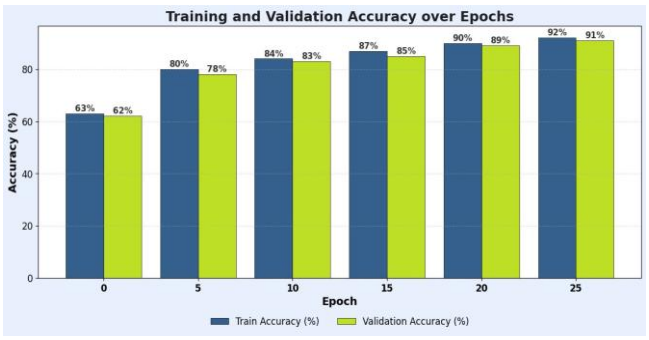
Epoch	Train Loss	Train Acc. (%)	Val. Loss	Val. Acc. (%)	ROC AUC
0	0.80	**63.00**	0.78	62.00	0.62
5	0.50	**80.00**	0.48	78.00	0.79
10	0.35	**84.00**	0.36	83.00	0.85
15	0.30	**87.00**	0.31	85.00	0.87
20	0.25	**90.00**	0.28	89.00	0.89
25	0.21	**92.00**	0.26	91.00	**0.91**



(a) ROC AUC over training epochs



(b) Loss of training and validation across epochs



(c) Epoch-wise training and validation cccuracy

Figure 3. Training performance metrics: ROC AUC, loss and accuracy.

4.2 Epoch-wise Evaluation of Training Dynamics

The performance of the suggested model was assessed for progressive learning behaviour through training and validation metrics over epochs. Table 3 reports the performance during each epoch in terms of loss, accuracy, and ROC AUC. Likewise, Figure 3 plots the training metrics from its evolution with time. At the initial epoch (Epoch 0), the model had limited predictive capacity; training accuracy of 63%, validation accuracy of 62%, and ROC AUC equal to 0.62. The network is untrained, as indicated by high loss values of 0.80 and 0.78 as part of this baseline performance. However, as training progressed, several things improved. By epoch 5, the validation accuracy was up to 78% while the ROC AUC improved sharply to 0.79. The model continues to improve performance with more epochs. The validation loss amounted to (0.31) with accuracy (85%) and AUC-ROC score (0.87) at epoch 15. Generalization has increased, and overfitting has decreased. Epoch 25 exhibits optimal performance, with a training accuracy of 92%, a validation accuracy of 91%, and a ROC AUC of 0.91. The model is capable of minimizing classification error, which leads to stable generalization performance. This corroborates the numerical findings, as illustrated in Fig 3. The training and validation sub-set loss curves exhibit a consistent fall, signifying smooth

convergence. Also, the accuracy plots indicate that both training accuracy and validation accuracy have almost the same upward trend. The ROC AUC curve further affirms that the model is tuning itself with every epoch to better discriminate between the positive and negative classes. In summary, the epoch-wise performance metrics affirm the robustness and convergence of the model. The steady improvement across loss, accuracy, and ROC AUC validates the effectiveness of the learning strategy and the suitability of the selected architecture for the classification task.

Table 4. Performance comparison of HWR-PDNet with existing models on the test dataset.

Model	**Acc.**	**Pre.**	**Re.**	**Spe.**	**F1**	**ROC AUC**
CNN (Baseline)	0.90	0.76	0.91	0.83	0.83	0.88
LSTM Model	0.89	0.78	0.87	0.80	0.82	0.86
CNN–Transformer	0.85	0.73	0.84	0.78	0.78	0.84
1D/2D/3D CNN	0.82	0.71	0.80	0.76	0.75	0.82
Beta-Elliptic + Fuzzy PD Classifier	0.88	0.79	0.86	0.81	0.82	0.87
Hybrid CNN–GRU Handwriting Classifies	0.90	0.80	0.89	0.85	0.84	0.89
Proposed HW R-PDNet	**0.92**	**0.81**	**0.95**	**0.89**	**0.88**	**0.91**

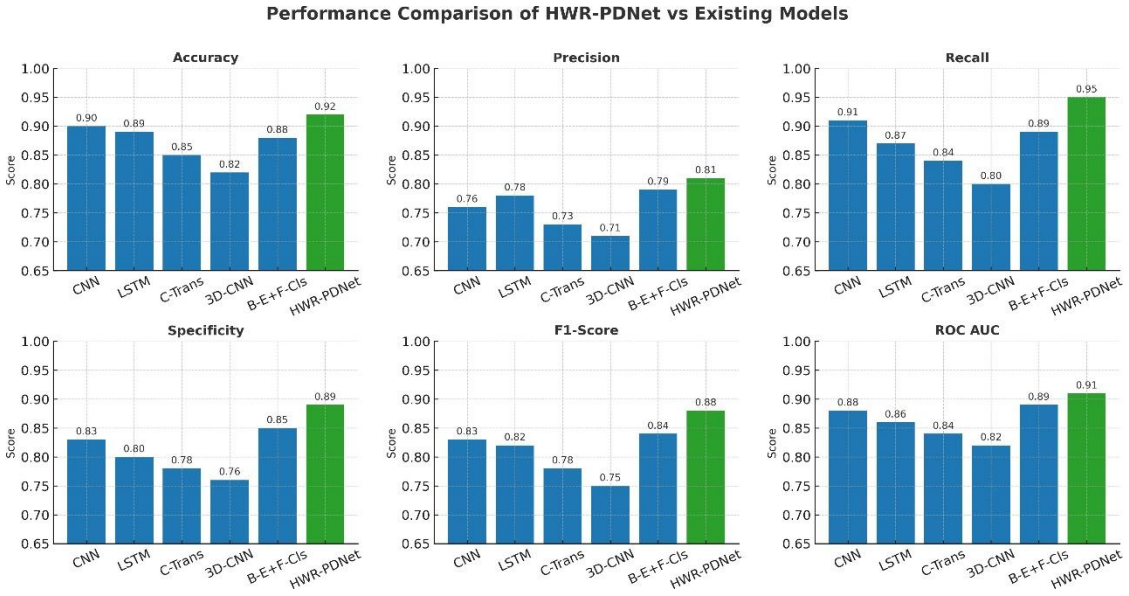


Figure 4. Performance comparison of HWR-PDNet *vs.* existing models' performance on the test dataset.

4.3 Comparative Examination with Current Models

Table 4 presents a detailed evaluation of the proposed HWR-PDNet framework against several contemporary baseline and hybrid models, including CNN (Baseline), LSTM Model, CNN–Transformer, 1D/2D/3D CNN, Beta-Elliptic + Fuzzy PD Classifier, and Hybrid CNN–GRU Handwriting Classifier. The baseline CNN achieved a strong recall of 0.91, but comparatively lower precision (0.76), suggesting a higher tendency toward false positives. The LSTM model showed balanced precision (0.78) and recall (0.87), though its overall accuracy (0.89) and ROC AUC (0.86) were slightly lower. The CNN–CNN-Transformer and 1D/2D/3D CNN architectures exhibited reduced performance, particularly in specificity, indicating limitations in correctly identifying healthy subjects. The Beta-Elliptic + Fuzzy PD Classifier demonstrated competitive precision (0.79) and specificity (0.81), while the Hybrid CNN–GRU Handwriting Classifier improved both accuracy (0.90) and specificity (0.85) compared to earlier baselines. In contrast, the proposed HWR-PDNet surpassed all other models, achieving the highest accuracy (0.92) and recall (0.95), alongside a robust F1-score (0.88) and the highest ROC AUC (0.91). Its specificity of 0.89 reflects an effective reduction in false positives, which is crucial in medical-screening applications. The graphical illustration in Figure 4 visually reinforces these results, showing HWR-PDNet's consistent lead across all metrics. This

performance gain can be attributed to its hybrid feature-extraction design, optimized regularization, and fine-tuning strategies, which enhance its generalization and discrimination capabilities. These findings confirm that HWR-PDNet is a reliable and superior choice for handwriting-based Parkinson's Disease detection in practical clinical workflows.

Table 5. Performance contribution of individual enhancements in the HWR-PDNet pipeline.

Configuration	Accuracy	F1-Score	ROC AUC
Baseline CNN (No Aug, No Fine-Tune)	0.86	0.82	0.84
With Grayscale Augmentation	0.88	0.84	0.86
With HSL Color Space Augmentation	0.89	0.86	0.88
With Dropout Regularization (p=0.2)	0.90	0.87	0.89
With Fine-tuning with Low LR	0.91	0.88	0.90
Full Model (HW R-PDNet)	**0.92**	**0.88**	**0.91**

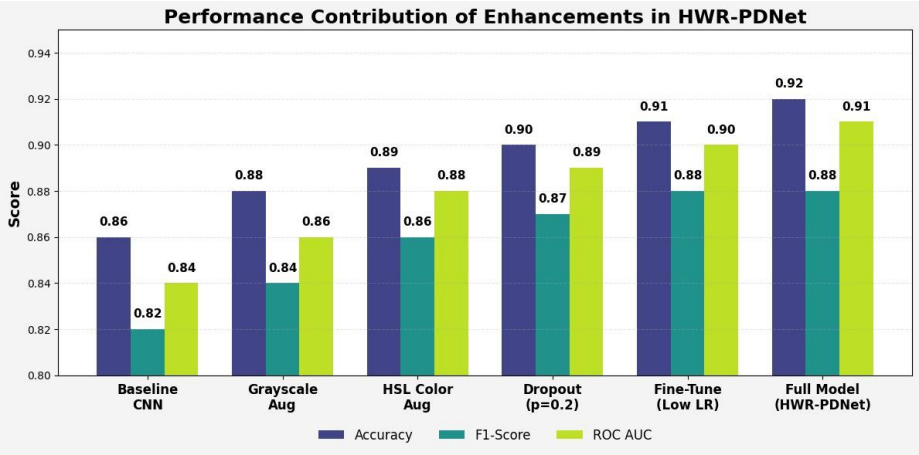


Figure 5. Performance contribution of enhancements in HWR-PDNet.

4.4 Impact Analysis of Incremental Enhancements in HWR-PDNet

An ablation study was conducted to elucidate the impact of specific enhancements in the HWR-PDNet architecture. As represented in Table 5, the classification metrics show progressive improvements with each enhancement, and the cumulative effects are visually summarized in Figure 5. Importantly, the configurations in Table 5 are cumulative rather than singular. Each successive configuration builds upon the enhancements of the previous one in the following order: baseline CNN without augmentation or fine-tuning, addition of grayscale augmentation, addition of HSL color space-augmentation (in addition to Grayscale), integration of dropout regularization with a probability of 0.2 (in addition to grayscale and HSL), fine-tuning with a low learning rate (in addition to grayscale, HSL, and dropout), and finally, the complete HWR-PDNet model that incorporates all enhancements. The order of integration was deliberately chosen to first expand the diversity and richness of the input representations (grayscale and HSL augmentations), then introduce regularization to mitigate overfitting (dropout), and finally apply targeted optimization to adapt the pre-trained backbone to the handwriting domain (fine-tuning). This approach ensures that the model initially develops a broader and more representative feature space, improves robustness against noise and overfitting, and then benefits from specialized adaptation to the target domain without catastrophic forgetting. Starting from the baseline CNN without data augmentation or fine-tuning, the model achieved 86% accuracy, an F1-score of 0.82, and ROC AUC of 0.84. Adding grayscale augmentation improved performance by enhancing the network's ability to detect contrast-based stroke patterns, leading to better generalization. Incorporating HSL color-space augmentation further increased accuracy to 89% and the F1-score to 0.86, showing the benefits of color-space diversity in capturing subtle handwriting variations. Integrating dropout regularization raised accuracy to 90% and ROC AUC to 0.89, demonstrating improved robustness. Fine-tuning with a low learning rate allowed the network to adapt feature representations more precisely to domain-specific characteristics, increasing accuracy to 91%

and F1-score to 0.88. The complete HWR-PDNet model, integrating all enhancements, achieved the best results: 92% accuracy, F1-score of 0.88, and ROC AUC of 0.91. These results confirm that the cumulative addition of augmentations, regularization, and fine-tuning significantly enhances both the generalization capability and the discriminative power of the proposed model.

Table 6. Per-class performance metrics stratified by stage.

Stage	Pre.	Re.	F1.	Support
Stage 1 (Early)	0.85	0.91	0.88	10
Stage 2 (Mild)	0.86	0.94	0.90	15
Stage 3 (Moderate)	0.87	0.93	0.90	13
Stage 4 (Severe)	0.80	0.89	0.84	10
Stage 5 (Advanced)	0.79	0.87	0.83	12
Healthy Controls	0.89	0.86	0.87	60

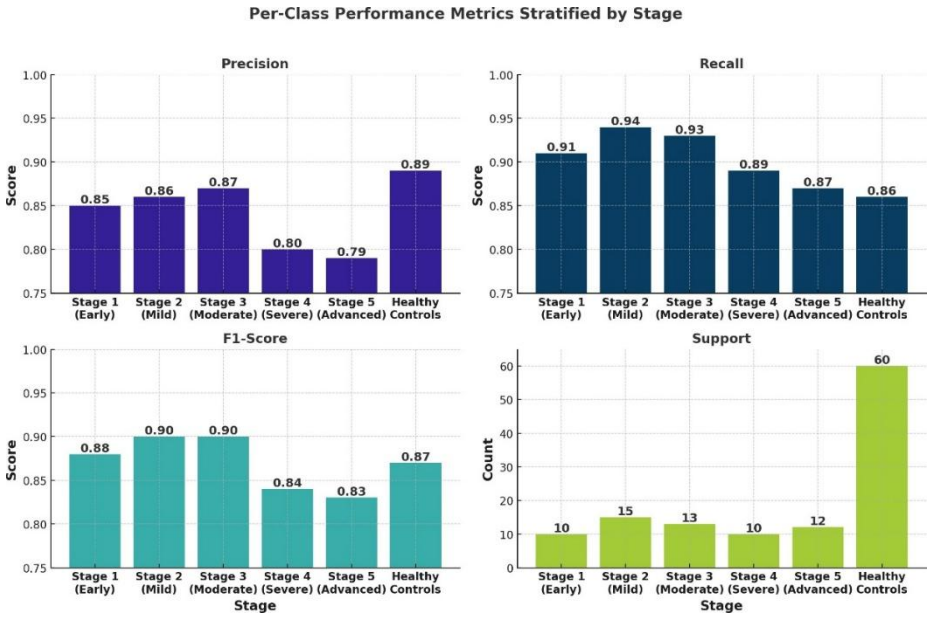


Figure 6. Per-class performance metrics stratified by stage.

4.5 Stage-wise Evaluation of Classification Performance

An analysis of the performance of the suggested HWR-PDNet across the various stages of PD was conducted to evaluate its discriminative capability, as represented in Table 6. The model demonstrated strong performance in detecting Stage 1 (Early) with a precision of 0.85, a recall of 0.91, and an F1-score of 0.88, indicating sensitivity to subtle handwriting irregularities associated with early neurodegeneration. Stages 2 (Mild) and 3 (Moderate) achieved the highest F1-scores of 0.90, supported by reliable precision–recall pairs, highlighting the model's robustness in capturing progressive motor impairments. Performance decreased slightly for Stage 4 (Severe) and Stage 5 (Advanced), with F1-scores of 0.84 and 0.83, respectively. This reduction can be attributed to overlapping clinical signs and reduced handwriting variability in advanced PD stages, though the model maintained consistent classification capability. The largest support was observed in the Healthy Control group (n=60), where the model reached an F1-score of 0.87. Notably, precision exceeded recall in this class (0.89 vs. 0.86), suggesting a conservative, but accurate, identification of healthy subjects. As illustrated in Figure 6, performance was relatively balanced across classes. Overall, the stage-wise evaluation underscores the ability of HWR-PDNet to effectively track disease progression, achieving higher efficiency in moderate-to-severe phases while preserving sensitivity at the early stage.

4.6 Discussion

The experimental evaluation of the proposed HWR-PDNet framework demonstrates consistent improvements across multiple classification metrics and experimental configurations. The epoch-wise

training dynamics (Table 3 and Figure 3) reveal a smooth convergence pattern with increasing accuracy and ROC AUC, indicating effective learning and generalization. Comparative analysis (Table 4 and Figure 4) shows that HWR-PDNet outperforms conventional CNN, LSTM, and hybrid models, achieving superior accuracy (92%) and recall (95%). The ablation study (Table 5 and Figure 5) highlights the cumulative contribution of augmentation, regularization, and fine-tuning, where each enhancement incrementally boosts model performance. Moreover, the stage-wise stratification (Table 6 and Figure 6) illustrates robust classification across all Parkinson's stages, with particularly strong results in early and moderate stages—underscoring the model's utility in early intervention scenarios. Collectively, the results confirm that HWR-PDNet offers a reliable, interpretable, and high-performing solution for stage-aware Parkinson's Disease recognition from handwriting data.

5. CONCLUSION

This study introduced HWR-PDNet, a hybrid deep-learning architecture for stage-specific classification of Parkinson's Disease (PD) using handwriting patterns. The model integrates convolutional feature extraction, attention-based enhancement, grayscale and HSL augmentations, dropout regularization, and fine-tuning with a low learning rate to achieve robust and generalizable outcomes. Experimental evaluations demonstrated that HWR-PDNet achieved superior classification performance compared to baseline models, with an overall accuracy of 92%, precision of 0.81, recall of 0.95, F1-score of 0.88, and ROC AUC of 0.91. The proposed framework consistently outperformed the baseline CNN (accuracy: 90%, F1-score: 0.83, ROC AUC: 0.88), LSTM (accuracy: 89%), and CNN–Transformer (accuracy: 85%) across all metrics. Ablation analysis confirmed the incremental contribution of each enhancement, with performance improving from 86% accuracy (baseline) to 92% in the final configuration. Stage-wise evaluation further highlighted the model's discriminative capacity: early-stage PD (Stage 1) achieved an F1-score of 0.88, moderate-stage PD (Stage 3) reached 0.90, and healthy controls were identified with an F1-score of 0.87, indicating low false-positive rates. Slightly lower scores were observed in advanced stages (Stages 4–5), reflecting the overlapping handwriting patterns typical of severe motor impairment. While this study focused on static handwriting images, the findings underscore the potential of extending HWR-PDNet to incorporate dynamic handwriting features, such as stroke velocity, acceleration, and temporal rhythm, which can be readily captured using touchscreen devices. Future work will explore the integration of such temporal signals with spatial handwriting patterns, along with multi-modal physiological data, to enable more sensitive, specific, and real-time PD monitoring. This direction holds promise for scalable, non-invasive, and personalized early intervention strategies in clinical and home settings.

REFERENCES

[1] M. Alissa et al., "Parkinson's Disease Diagnosis Using Convolutional Neural Networks and Figure-copying Tasks," Neural Computing and Applications, vol. 34, no. 2, pp. 1433–53, Sept. 2021.

[2] M. Hireš et al., "Convolutional Neural Network Ensemble for Parkinson's Disease Detection from Voice Recordings," Computers in Biology and Medicine, vol. 141, pp. 105021–105021, Nov. 2021.

[3] H. Chen et al., "A Parkinson's Disease-related Nuclei Segmentation Network Based on CNN-Transformer Interleaved Encoder with Feature Fusion," Computerized Medical Imaging and Graphics, vol. 118, p. 102465, 2024.

[4] N. Aggarwal, B. S. Saini and Savita Gupta, "A Deep 1-D CNN Learning Approach with Data Augmentation for Classification of Parkinson's Disease and Scans without Evidence of Dopamine Deficit (SWEDD)," Biomedical Signal Processing and Control, vol. 91, p. 106008, 2024.

[5] X.-C. Wang et al., "Comparison of One- Two- and Three-dimensional CNN Models for Drawing Test-based Diagnostics of Parkinson's Disease," Biomedical Signal Processing and Control, vol. 87, p. 105436, 2024.

[6] M. F. Allebawi et al., "Parkinson's Disease Detection from Online Handwriting Based on Beta-Elliptical Approach and Fuzzy Perceptual Detector," IEEE Access, vol. 12, pp. 56936-56950, 2024.

[7] L. Sigcha, L. Borz and G. Olmo, "Deep Learning Algorithms for Detecting Freezing of Gait in Parkinson's Disease: A Cross-dataset Study," Expert Systems with Applications, vol. 255, Part A, p. 124522, 2024.

[8] G. Celik and E. Başaran, "Proposing a New Approach Based on Convolutional Neural Networks and Random Forest for the Diagnosis of Parkinson's Disease from Speech Signals," Applied Acoustics, vol. 211, p. 109476, 2023.

[9] N. Madusanka and B.-il Lee, "Vocal Biomarkers for Parkinson's Disease Classification Using Audio

Spectrogram Transformers," Journal of Voice, in Press, 10.1016/j.jvoice.2024.11.008, 2024.
[10] N. Khalid and M. S. Ehsan, "Critical Analysis of Parkinson's Disease Detection Using EEG Sub-bands and Gated Recurrent Unit," Engineering Science and Technology, an Int. J., vol. 59, p. 101855, 2024.
[11] Z. Li et al., "A Survey of Convolutional Neural Networks: Analysis, Applications and Prospects," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 6999-7019, 2021.
[12] Y. Qi et al., "A Comprehensive Overview of Image Enhancement Techniques," Archives of Computational Methods in Engineering, vol. 29, pp. 583-607, 2021.
[13] D. A. Van Dyk and X.-L. Meng, "The Art of Data Augmentation," Journal of Computational and Graphical Statistics, vol. 10, no. 1, pp. 1-50, 2001.
[14] D. Ping Tian, "A Review on Image Feature Extraction and Representation Techniques," Int. Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 4, pp. 385-396, 2013.
[15] K. Dong et al., "MobileNetV2 Model for Image Classification," Proc. of the 2020 2nd IEEE Int. Conf. on Information Technology and Computer Application (ITCA), pp. 476-480, Guangzhou, China, 2020.
[16] J. He et al., "ReLU Deep Neural Networks and Linear Finite Elements," arXiv preprint arXiv:1807.03973, 2018.
[17] Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," Proc. of the 2018 IEEE/ACM 26th Int. Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 2018.
[18] F. Zhuang et al., "A Comprehensive Survey on Transfer Learning," Proceedings of the IEEE, vol. 109, no. 1, pp. 43-76, 2020.
[19] F. Radenović, G. Tolias and O. Chum, "CNN Image Retrieval Learns from BoW: Unsupervised Fine-tuning with Hard Examples," Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016.
[20] C. S. Corley et al., "Exploring the Use of Deep Learning for Feature Location," Proc. of the 2015 IEEE Int. Conf. on Software Maintenance and Evolution (ICSME), Bremen, Germany, pp. 556-560, 2015.
[21] X. Jiang, H. Yu, J. Yang, X. Liu and Z. Li, "A New Network Structure for Parkinson's Handwriting Image Recognition," Medical Engineering & Physics, vol. 139, Article no. 104333, 2025.
[22] H. Lu, G. Qi, D. Wu et al., "A Novel Feature Extraction Method Based on Dynamic Handwriting for Parkinson's Disease Detection," PloS One, vol. 20, no. 1, Article no. e0318021, 2025.
[23] I. Kansizoglou, K. A. Tsintotas, D. Bratanov and A. Gasteratos, "Drawing-aware Parkinson's Disease Detection through Hierarchical Deep Learning Models," IEEE Access, vol. 13, pp. 21880-21890, 2025.
[24] A. S. M. Miah, T. Suzuki and J. Shin, "A Methodological and Structural Review of Parkinson's Disease Detection across Diverse Data Modalities," IEEE Access, vol.13, pp. 98931-98975, 2025.
[25] A. Javeed, L. Ali, R. Nour, A. Noor and N. Golilarz, "Handwriting-based Detection of Schizophrenia and Bipolar Disorder Using Machine Learning," IEEE Sensors J., vol. 25, no. 5, pp. 9113-9120, 2025.
[26] A. Al-Shannaq and L. Elrefaei, "Age Estimation Using Specific Domain Transfer Learning," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 6, no. 2, pp. 122–139, 2020.

ملخص البحث:

مرض باركنسون عبارة عن خللٍ عصبي مزمن ومتقدّم يتميز بخصائص غير اعتيادية في النّظام الحركي، ويمكن الكشْف عن الحالة في مراحلها المبكّرة عن طريق عدم انتظام الكتابة اليدوية للفرد. ويُعدّ التّشخيص المبكّر أمراً بالغ الأهمية؛ فهو يمكّن من التّدخّل العلاجي في الوقت المناسب، إضافةً إلى أنّه يبين مرحلة تقدُّم المرض. ومع ذلك تُعاني الطُّرق التّقليدية في التّشخيص من محدِّداتٍ تتعلّق بإمكانية التّوسع وانخفاض الحساسية.

تقترح هذه الورقة بنية تعلُّم عميق تستخدم الكتابة اليدوية بهدف الكشف المبكّر عن مرض باركنسون. وقد اشتملت مجموعة البيانات المستخدمة في هذه الدراسة على 816 عينة كتابة يدوية تعود لـِ 120 شخصاً. وتجدر الإشارة إلى أنّ النّظام المقترح حقّق مؤشرات أداء جيدة تفوق مؤشرات الأداء لعددٍ من الطُّرق التّقليدية من حيث المتانة والحساسية. ويُعدّ إطار العمل الذي تنطوي عليه الطّريقة المقترحة إطاراً خفيف الوزن مُناسباً لتطبيقات الكشْف عن المرض في الزّمن الحقيقي، وهو يوفّر إمكانية مهمّة لدعم القرار السّريري وتطوّر الكشْف عن الإصابة بالمرض من عدمها عن بُعد.