Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 1 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i01.58513

PAPER

Parkinson's Disease Detection through Offline Handwriting Analysis: A CNN-Based Approach

Ameur Bensefia[1] (✉), **Chawki Djeddi**[2,3], **Abdelhakim Hannousse**[4], **Moizes Diaz**[5]

[1]Higher Colleges of Technology, Abu Dhabi, UAE

[2]Université Larbi Tebessi, Tébessa, Algeria

[3]Université de Rouen-Normandie, Rouen, France

[4]Université 8 Mai 1945, Guelma, Algeria

[5]Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

abensefia@hct.ac.ae

ABSTRACT

This study investigates the use of offline handwriting analysis for the automated diagnosis of Parkinson's disease (PD) using deep learning techniques. A convolutional neural network (CNN) was designed to extract discriminative spatial features from handwritten images and classify subjects as either PD patients or healthy controls. The model was evaluated on four publicly available datasets—HandPD, NewHandPD, PaHaW, and UCI—representing a diverse range of handwriting patterns and acquisition conditions. The proposed CNN achieved 100% accuracy on the smaller UCI dataset and 94.74% accuracy on the larger NewHandPD dataset. To overcome dataset imbalance and limited sample diversity, various data augmentation strategies were applied, leading to a notable increase in overall performance, with accuracies exceeding 97% on larger datasets. These results demonstrate that offline handwriting analysis, supported by deep CNN architectures and data augmentation, offers a promising, non-invasive, and cost-effective approach for early PD diagnosis and potential continuous monitoring. Furthermore, this study aligns with broader advances in AI-assisted medical diagnostics, reinforcing the role of machine learning and image-based analysis in healthcare applications.

KEYWORDS

parkinson's disease (PD), convolutional neural network (CNN), handwriting, archimedean spiral

1 INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by motor symptoms such as tremors, bradykinesia, rigidity, and postural instability. Alongside these motor symptoms, non-motor symptoms, including cognitive impairment, mood disturbances, and autonomic dysfunction, contribute to the multifaceted nature of the disease [1]. Accurate diagnosis of PD is crucial for timely intervention and optimal management of symptoms. Traditional diagnostic approaches for PD primarily rely on clinical assessment, neuroimaging techniques,

Bensefia, A., Djeddi, C., Hannousse, A., Diaz, M. (2026). Parkinson's Disease Detection through Offline Handwriting Analysis: A CNN-Based Approach. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(1), pp. 133–146. https://doi.org/10.3991/ijoe.v22i01.58513

Article submitted 2025-09-03. Revision uploaded 2025-10-21. Final acceptance 2025-10-21.

and biochemical markers. While effective, these methods often lack sensitivity and specificity, particularly in the early stages of the disease. Moreover, they can be costly and resource-intensive, which limits widespread adoption.

Recent advancements in digital health technologies have opened avenues for exploring alternative diagnostic modalities that offer non-invasive, cost-effective, and accessible solutions for PD diagnosis. Among these modalities, offline handwriting analysis shows promise in detecting subtle motor and cognitive impairments associated with PD [2]. Handwriting, a complex motor task, involves coordinated movements of the hand, fingers, and wrist, as well as cognitive processes such as planning, attention, and memory. Studies have demonstrated that individuals with PD exhibit distinctive handwriting features, including micrographia (reduced letter size), dysfluencies (hesitations and interruptions), and alterations in stroke dynamics [3]. These alterations reflect underlying motor dysfunction, bradykinesia, and impaired fine motor control characteristic of Parkinson's disease.

Hand tremors are one of the primary motor symptoms associated with PD, sparking considerable interest in handwriting analysis as a potential diagnostic tool. This approach involves examining hand movements during handwriting, either through online trace analysis or by capturing hand movements using smart pens. By leveraging digital pen technology and advanced handwriting analysis algorithms, researchers can precisely measure and quantify subtle abnormalities in handwriting, such as variations in pen pressure, velocity, stroke duration, and spatial-temporal parameters [8]. Despite the acknowledged benefits of handwriting analysis, including its non-invasive nature and cost-effectiveness, the literature has not extensively explored the use of offline handwriting analysis for diagnosing PD. Offline analysis offers distinct advantages, such as the ability to detect and monitor the disease economically and without the need for continuous observation. Given that PD impacts motor control and may lead to noticeable changes in handwriting, analyzing handwriting samples provides a practical approach to assessing motor symptoms. This study aims to fill the existing gap in the literature by focusing on offline handwriting analysis, particularly static methods that involve analyzing handwriting images. We conducted extensive experiments using several publicly available datasets to demonstrate the effectiveness of these approaches. To this end, we developed a CNN model architecture to evaluate its performance in detecting PD across different datasets, thereby advancing the understanding of how static handwriting analysis can contribute to diagnosis and monitoring.

This paper is organized into four sections as follows: In Section 2, we review advancements in PD diagnosis through both offline and online handwriting. Next, we detail our proposed approach based on a CNN network in Section 3, followed by the presentation of experimental findings in Section 4. Finally, we discuss future research directions in this area.

2 LITERATURE REVIEW

Several works have been dedicated to the detection of PD through the analysis of handwriting images [4] [5]. These studies typically fall into two primary categories: static and dynamic approaches. Static approaches involve the examination of static features such as the shape, size, and spatial distribution of handwritten characters, while dynamic approaches focus on analyzing real-time captured data such as pen pressure, velocity, and acceleration during the act of writing. In the following, we delineate each approach and highlight key studies from the literature that exemplify these categories.

2.1 Static approaches

Static approaches involve the analysis of images of handwriting samples, such as scanned documents or digital images. These methods extract features concerning the shape, size, and consistency of handwritten characters and symbols. Techniques such as deep learning, transfer learning, and traditional image processing are used to classify these features and identify potential markers of PD. In the following section, we present some of the major works that have adopted this approach.

Mitra et al. [6] fine-tuned a pre-trained ResNet-152 model for PD detection by freezing its top 18 layers to retain the learned features from the original training. They further enhanced the training process by incorporating callbacks and applying early stopping to prevent overfitting and ensure optimal training duration. The model was experimented on the NewHandPD dataset, which was augmented with horizontal flips to increase the diversity and robustness of the training samples. This fine-tuning and augmentation strategy led to the model achieving an impressive 100% accuracy on the dataset.

Agrawal et al. [7] proposed a hybrid deep learning and machine learning technique for PD detection. They used pretrained CNN models on the NewHandPD dataset, with VGG16 performing best. Features from VGG16's fc8 layer were optimized using Binary Grey Wolf Optimization (BGWO). The selected features were classified using an SVM. This approach achieved 99.8% accuracy on the NewHandPD dataset. The results highlight the effectiveness of combining deep learning and traditional ML techniques.

Wang et al. [8] proposed a weighted voting method using logistic regression (LR), decision trees (DT), and K-nearest neighbors (KNN). These classifiers utilized fused features from two transformers: Vision Transformer (ViT) and Coordinate Attention-enhanced Swin Transformer (CAS). The HandPD and NewHandPD datasets were augmented using CycleGAN to generate synthetic images, improving robustness. Fused features captured both global and local dependencies, creating a comprehensive feature set. The weighted voting method assigned different weights to LR, DT, and KNN based on their performance. This approach achieved 92.68% accuracy on the augmented datasets. It highlights the benefits of advanced feature extraction, classifier fusion, and sophisticated data augmentation.

Kamran et al. [9] evaluated CNN variants for PD handwriting classification using four datasets: PaHaW, HandPD, NewHandPD, and Parkinson's Drawing. PaHaW signals were converted to RGB images for CNN compatibility. To address data limitations, they applied augmentation techniques, including contrast adjustment, illumination changes, thresholding, flipping, and rotation. Six CNN architectures were tested: 1) AlexNet, 2) GoogleNet, 3) VGG16, 4) VGG19, 5) ResNet50, and 6) ResNet101. The highest accuracy (99.22%) was achieved using AlexNet on the combined dataset with illumination augmentation, while performance on HandPD alone reached 90.41%.

Naz et al. [10] investigated CNN-based PD detection with feature fusion using AlexNet, GoogleNet, VGG16, VGG19, ResNet50, and ResNet101 on HandPD, NewHandPD, and Parkinson's Drawing datasets. Augmentation techniques (rotation, flipping, contrast, and illumination) enhanced the dataset. Pretrained CNN features were extracted with frozen layers, fused via addition, multiplication, and mean operations, and classified using SVM. The best accuracy (99.35%) was obtained by fusing features from the fc6 layer of AlexNet and VGG16, using illumination-based augmentation.

2.2 Dynamic approaches

Dynamic approaches focus on the analysis of handwriting data collected in real-time using digital devices equipped with sensors. These methods capture dynamic

features such as pen pressure, velocity, and acceleration, providing detailed information about the motor control involved in handwriting. This data is often analyzed using traditional and advanced machine learning techniques to detect patterns indicative of Parkinson's disease.

Valla et al. in [11] have introduced an innovative feature derived from Archimedean spiral drawing tests for diagnosing PD using traditional machine learning techniques. The features captured subtle changes in handwriting trajectory that are difficult to discern visually but may be indicative of tremor-like symptoms. The study utilizes two datasets, DraWritePD and PaHaW, employing Fisher's score and recursive feature elimination for feature selection. Six classifiers were trained and evaluated through nested cross-validation to distinguish between healthy controls and Parkinson's patients. The ensemble classifiers, combined with a nested wrapper-type feature selection method, achieve an accuracy of 84.33% and 73.71% for DraWritePD and PaHaW datasets, respectively.

Diaz et al. [12] introduced "dynamically enhanced" static handwriting images for PD diagnosis, preserving temporal and velocity data by plotting points with pen-ups. Tested on the PaHaW dataset, this method outperformed static and dynamic handwriting used separately. Diaz et al. [12] later analyzed handwriting dynamics, including pen pressure and stroke speed, using online acquisition tools. Their model, combining 1D convolutions and BiGRUs, excelled in detecting Parkinsonian symptoms, achieving superior results on PaHaW and competitive performance on NewHandPD.

Lamba et al. [13] utilized the UCI PD Spiral Drawings dataset to extract 29 kinematic features from the time-series data for both static and dynamic tests. These features included metrics such as the number of strokes, speed, rate of change of displacement, rate of change of acceleration with respect to time, number of changes in velocity, and the total time the pen is in the air during the test. To address the imbalanced nature of the dataset, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. For feature selection, both genetic algorithms and mutual information gain methods were applied. The best performance was achieved by selecting nine features using the mutual information gain method. This approach resulted in a perfect accuracy rate of 100% and an F1 score of 95.79%.

Xu et al. [14] proposed a majority voting approach involving six Random Forest (RF) models, each trained separately using different sensor signals from the NewHandPD dataset. Each of the six handwritten sensor signals was subsampled into segments with a channel length of N = 3000. To address the imbalanced nature of the dataset, the authors used stratified five-fold cross-validation. This method ensures that each fold maintains the same proportion of class labels as the original dataset, providing a more reliable evaluation of the model's performance. By employing a majority voting scheme across the six RF models, the authors achieved an accuracy of 89.40%.

Drotar et al. [15] enhanced PD detection by expanding their handwriting analysis feature set. Original kinematic features included speed, NCV/NCA, writing duration, stroke velocity, acceleration, jerk, and stroke dimensions [16]. New features introduced entropy (Shannon, Renyi) and energy measures (CE, TKE) to assess handwriting irregularities. Empirical mode decomposition (EMD) extracted intrinsic data, and an SVM with a radial Gaussian kernel classified 39 PD and 38 healthy subjects. The best single feature achieved 76% accuracy, while combining 168 features improved accuracy to 85.6%.

Table 1 summarizes the discussed works, highlighting the datasets used, the types of features employed (static or dynamic), and the highest accuracy performances achieved.

2.3 Related Works in Medical Image Analysis Using AI

While handwriting analysis for PD detection represents a specific domain of biomedical pattern recognition, related studies in other medical contexts have demonstrated the versatility and effectiveness of AI and deep learning in disease diagnosis. For instance, Al-Nawashi et al. [17] proposed a machine learning framework for breast cancer detection, integrating multiple learning models to enhance diagnostic accuracy. Similarly, Gharaibeh et al. [18] introduced a Swin Transformer-based segmentation and multi-scale feature pyramid fusion module for Alzheimer's disease detection, achieving robust feature extraction and improved classification results. Moreover, Al-Hazaimeh et al. [19] combined artificial intelligence and image processing for diagnosing diabetic retinopathy from retinal fundus images, demonstrating the capacity of AI to identify subtle pathological patterns in medical imagery.

In addition, Abu-Amara et al. [20] explored robot-based therapy for improving academic skills in children with autism, illustrating how intelligent robotic systems can support cognitive and motor development through adaptive interaction. Furthermore, the authors in [21] integrated robotic kinematics and dynamics with online handwriting features for dysgraphia classification, highlighting the potential of combining biomechanical and spatiotemporal data with deep learning to improve diagnostic precision.

These works collectively reinforce the growing evidence that AI-driven methods—especially those leveraging transformer architectures and hybrid ML pipelines—can effectively capture complex visual and biological patterns across diverse medical imaging applications. This cross-domain success further validates the use of deep learning and transformer-based models for handwriting-based PD detection.

Table 1. Comparison of PD approaches and performances

	Dataset	Approach	Performance Rate
Diaz [12]	PaHaW	Static & Dynamic	86.67%
Mitra [6]	NewHandPD	Static	100%
Agrawal [7]	NewHandPD	Static	99.8%
Wang [8]	HandPD NewHandPD	Static	88.92% 92.68%
Kamran [9]	PaHaW HandPD NewHandPD Combined datasets with augmentation	Static	62.50% 90.41% 98.31% 99.22%
Naz [10]	HandPD NewHandPD	Static	99.35%
Valla [11]	DraWrite PaHaW	Dynamic	84.33% 73.71%
Diaz [15]	PaHaW NewHandPD	Dynamic	93.75% 94.44%
Lamba [13]	UCI Dataset	Dynamic	96.02%
Xu [14]	NewHandPD	Dynamic	89.40%
Drotar [16]	PaHaW	Dynamic	88.13%

3 METHODOLOGY

One of the prominent motor symptoms associated with PD is hand tremors. These tremors can significantly impact hand control, leading to difficulties in performing fine motor tasks such as handwriting. Consequently, there is significant interest in exploring handwriting analysis for detecting this condition, as highlighted in existing literature. This involves analyzing hand movements during the process of handwriting acquisition, either through online trace analysis or by examining hand movements recorded by smart pens. However, despite the acknowledged benefits, such as non-invasiveness and cost-effectiveness, the literature has not extensively explored the use of offline handwriting analysis for diagnosing PD. Offline handwriting analysis offers several advantages, including providing a non-invasive and economical means for detection and monitoring of the disease. Given that PD affects motor control and may cause changes in handwriting, analyzing handwriting samples presents a convenient approach to assessing motor symptoms.

3.1 Datasets

Various datasets have been used for the automatic detection of PD through handwriting analysis. To evaluate our model, we used the four publicly available datasets, namely HandPD, NewHandPD, PaHaW, and UCI. The details and the preparation process of each dataset are described in the following sections:

HandPD dataset. HandPD is a dataset specifically designed for PD, comprising handwritten samples from two groups: a healthy control group and a PD patient group. The data were collected at Botucatu Medical School, São Paulo State University, Brazil [22].

The dataset includes 18 healthy individuals (six males and 12 females) aged between 19 and 79 years, and 74 PD patients (59 males and 15 females) aged between 38 and 78 years. In terms of handedness, the control group included 16 right-handed and two left-handed individuals, with a mean age of 44.22 ± 16.53 years. The patient group, on the other hand, comprised 69 right-handed and five left-handed individuals, demonstrating a mean age of 58.75 ± 7.51 years. To construct the dataset, each subject completes a series of tasks, including drawing circles, Archimedean spirals, and meanders on a prefilled form. In this study, we focus on analyzing the Archimedean spiral drawing.

Dataset Preparation. The handwritten traces (blue traces) were drawn on top of a predefined spiral template (black traces), where the two traces overlap and intertwine, which makes them unusable for our model, as shown in Figure 1; therefore, a cleaning phase was needed to separate the two traces. To this end, and to identify the blue trace within each image, we established a thresholding mechanism for each channel of the RGB color space as follows:

$$BlueTrace_I = \begin{cases} I_i \ \ if \ \ I_{R_i} \le T_R \ and \ I_{G_i} \le T_G \ and \ I_{B_i} \le T_B \\ \\ 255 \end{cases} \quad (1)$$

Where I_i represents the pixel's original intensity, R, G, and B represent the different intensities of the pixel i in the image I on the channels Red, Green, and Blue, respectively, and TR, TG, and TB represent the thresholds of the three channels set experimentally. Samples of the resulting images are illustrated in Figure 1.

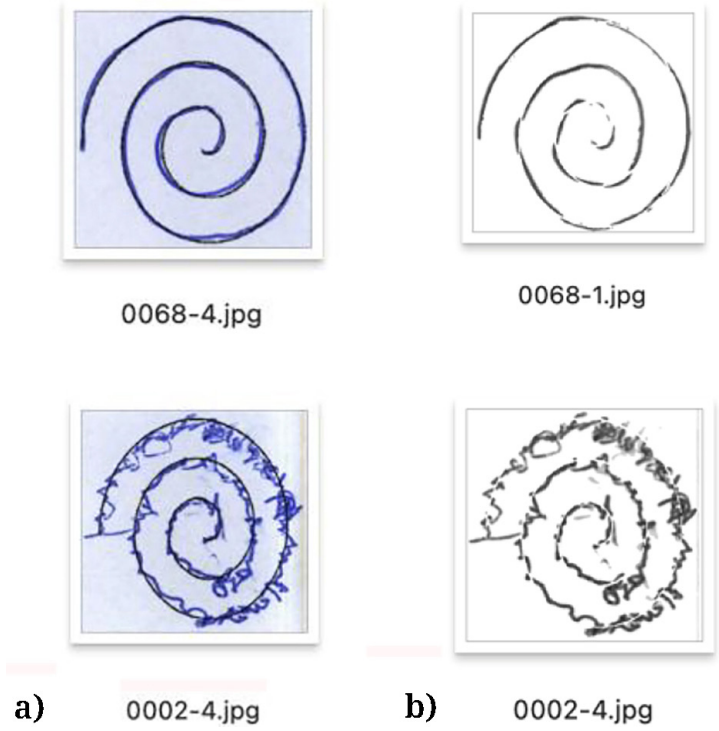

Fig. 1. HandPD Archimedean spiral drawing samples before (a) and after removing the blue lines

NewHandPD dataset. The improved HandPD dataset [23], known as NewHandPD, includes 66 individuals divided into two groups: healthy and patient. The healthy group consists of 35 individuals (18 male, 17 female) with an average age of 44 years, while the patient group includes 31 individuals (21 male, 10 female) with an average age of 58 years. Each participant completed 12 exams, generating nine images and 12 signals per individual, including spirals, meanders, circular movements, and diadochokinetic tasks. During the samples' collection, handwritten dynamics were recorded by means of a smart pen, which means we have images from spirals (4), meanders (4), circles on the paper (1), and signals for all 12 exams. In total, the dataset contains 264 images and 792 signals, offering a more balanced and comprehensive set of data than the original HandPD dataset.

In our study, for consistency purposes, we considered only the Archimedean spiral samples, which included 124 samples from PD subjects and 140 samples from healthy subjects. These samples were processed according to the procedure applied to the HandPD samples to separate the handwritten traces (blue traces) from the pre-defined spiral template. Ultimately, we obtained 64 exploitable samples from healthy subjects and 124 samples from PD subjects (see Figure 2a).

UCI dataset. This dataset was created at the Department of Neurology in the Cerrahpasa Faculty of Medicine, Istanbul University [24]. 62 people with PD and 15 healthy subjects were asked to create drawings using developed software for coordination tests. Three types of tests were developed for data collection using a graphics tablet. The first is the static spiral test (SST), commonly used in clinical research to assess motor performance, measure tremors, and diagnose PD. In this test, patients are required to retrace Archimedean spirals displayed on the tablet. Data, including the features mentioned, are recorded during the test. The second test is the dynamic spiral test (DST). Unlike the SST, the Archimedean spiral in this test blinks on and off

at intervals, requiring patients to remember the pattern. Most patients struggled to retain the pattern, resulting in poorer performance compared to the SST. The third test is the stability test on a certain point (STCP). In this test, subjects are asked to hold the digital pen on a red point in the center of the screen without touching it for a set duration. The aim is to assess hand stability or tremor levels.

For our experiments, we opted to use only the SST samples, as they were collected using a similar approach as the other datasets. As with the PaHaW dataset, the selected samples were processed to create binary images based on the x and y coordinates of each pen position (see Figure 2b).

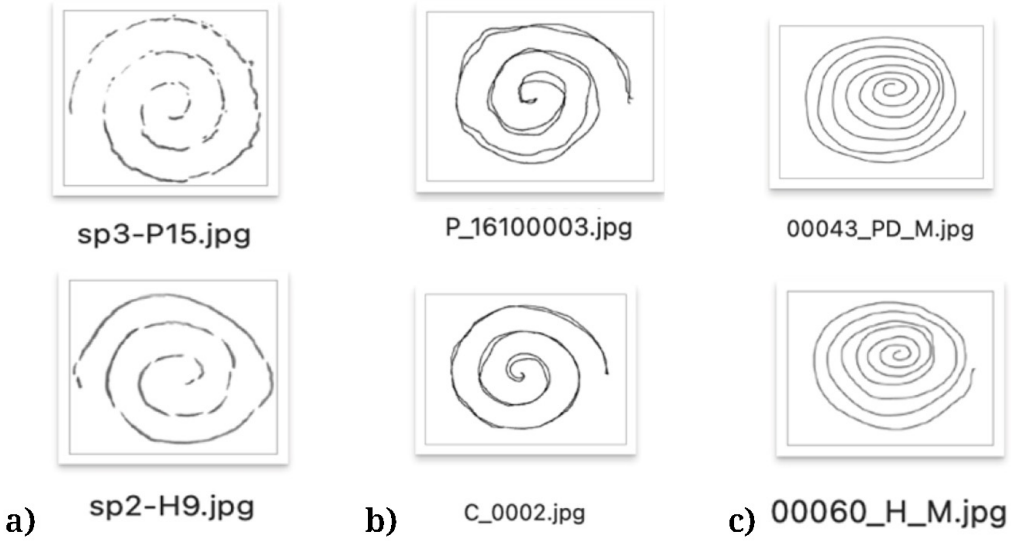

Fig. 2. Archimedean spiral drawing samples of healthy and patient subjects from (a) NewHandPD dataset, (b) UCI dataset, and (c) PaHaW dataset

PaHaW dataset. The PaHaW dataset was developed at the Department of Neurology, Masaryk University, and St. Anne's University Hospital in Brno, Czech Republic [25]. It includes data from 37 (PD) patients (19 men and 18 women) and 38 healthy control subjects (20 men and 18 women). All participants were asked to provide handwriting and drawing samples by following a template consisting of eight different tasks, including the Archimedean spiral drawing. These samples were captured using a tablet, with various features such as pen movement, coordinates, and pressure being recorded.

To achieve results comparable with other approaches, we focused solely on the Archimedean samples. These samples were processed to create binary images based on the x and y coordinates of each pen position (see Figure 2c).

3.2 Data augmentation

To address the issue of unbalanced datasets, with a disproportionate number of samples between healthy and patient subjects, as well as the challenge posed by datasets with relatively low sample sizes, we have implemented data augmentation techniques to increase the number of samples. These techniques include horizontal flips, vertical flips, scaling, and noise adjustment operations. These augmentations are beneficial to our system architecture described in the following section.

Table 2 summarizes the original number of samples for each dataset, as well as the final number of samples after augmentation.

Table 2. Datasets size before and after augmentation

Datasets	Original Dataset		Original Dataset with Augmentation	
	Healthy Subject	Patient Subject	Healthy Subject	Patient Subject
HandPD	72	296	360	1480
NewHandPD	64	124	320	620
PaHaW	36	36	180	180
UCI	15	25	75	125
All Datasets	187	481	935	2405

3.3 System architecture

We propose a neural network architecture that comprises two modules of layers: convolutional layers and dense layers, both designed to extract hierarchical features from the handwritten traces' images for classification either as PD trace or healthy subjects.

Convolutional neural networks (CNNs) are integral in computer vision for their ability to learn hierarchical patterns from images, crucial for tasks like image classification and object detection. By mimicking the human visual cortex, CNNs capture pixel-level relationships efficiently. Our deep learning model employs a CNN to analyze handwritten samples, focusing on detecting patterns in Archimedean spiral drawings to identify Parkinson's disease patients.

The model begins with an image input layer, accepting grayscale images of size 96 × 96 pixels. Subsequently, the first module of three convolutional layers is employed to convolve input feature maps with learnable filters, followed by rectified linear unit (ReLU) activation functions to introduce non-linearity. Since we need to preserve the tiny details in the input images, we began the convolution with a filter size of 7 × 7 followed by filters of 3 × 3.

Batch normalization layers are applied after each convolutional layer to stabilize training by normalizing the activations. Additionally, cross-channel normalization layers are utilized to enhance the response of specific neurons. Max pooling layers are then employed to down-sample feature maps, reducing spatial dimensions while preserving essential information.

The module of convolutional layers is followed by a module of dense layers made up of two fully connected layers, which serve as the classifier by learning high-level representations of the input features. Each fully connected layer is connected to all neurons in the previous layer, enabling complex feature combinations to be learned. Dropout layers are inserted after the fully connected layers to prevent overfitting by randomly dropping a fraction of the neurons during training. The final layer of the network is a softmax layer, which computes the probability distribution over the classes and facilitates multi-class classification. The softmax layer is coupled with a classification layer, which computes the loss and accuracy metrics during training.

The model is trained using the stochastic gradient descent with momentum (SGDM) optimization algorithm, with an initial learning rate of 0.001. The training process is executed for a maximum of 25 epochs, with data shuffled at every epoch.

Validation data is used to monitor the model's performance and prevent overfitting, with validation occurring every 30 iterations.

4 EXPERIMENTATION AND RESULTS

Although there is no universal agreement on the split ratio for datasets between training and testing, we have opted for an 80/20 (training/test) division, which is commonly used. In the first stage, the experiments were conducted on all the different datasets separately and then by combining them all into one large dataset. This scenario has been repeated for datasets with and without augmentation. For each dataset, we varied the number of samples used when applicable. Specifically, we tested our system with a balanced number of samples between healthy and patient subjects, as well as with an unbalanced number of samples.

To evaluate our model as accurately as possible, we ran the model five times, reshuffling the samples with each iteration, and reported the highest and average accuracy (including standard deviation). The loss scores for training and testing, as well as the specificity, sensitivity, and F1 scores for the maximum accuracy, were recorded for each scenario. The results of these experiments are summarized in Tables 3 and 4.

Table 3. Performance on the datasets without augmentation

Dataset	# Samples Healthy:PD	Accuracy			Loss		Confusion Matrix (TP, FN/FP, TN)	Sensitivity: Specificity	F1-Score
		Max	Mean	Std	Train	Test			
HandPD	72: 72 72: 296	90.41 94.52	87.95 92.05	0.022 0.017	0.l01 0.127	0.346 0.244	72, 0/8, 64 254, 42/2, 70	100.00:89.39 85.71:96.61	66.67 85.71
NewHandPD	64: 64 64: 124	92.31 94.74	83.03 88.95	0.079 0.039	0.140 0.232	0.238 0.157	64, 0/9, 55 124, 0/5, 59	100.00:86.67 100.00:92.59	91.67 91.67
PaHaW	36: 36	87.93	75.71	0.139	0.143	0.459	32, 4/0, 36	87.50:100.00	93.33
UCI	15: 15 15: 25	100 100	83.33 92.50	0.117 0.111	0.003 0.069	0.007 0.100	15, 0/0, 15 25, 0/0, 15	100.00:100.00 100.00:100.00	100.0 100.0
All Datasets	187:187 187:481	91.89 92.48	87.03 89.02	0.036 0.022	0.197 0.167	0.216 0.192	176, 11/19, 168 450, 31/15, 172	94.29:89.74 93.55:92.16	91.67 85.29

Table 4. Performance on the datasets with augmentation

Dataset	# Samples (Healthy:PD)	Accuracy			Loss		Confusion Matrix (TP, FN/FP, TN)	Sensitivity: Specificity	F1-Score
		Max	Mean	Std	Train	Test			
HandPD	360:360 360:1480	92.36 97.28	91.25 96.14	0.012 0.014	0.035 0.029	0.281 0.085	323, 37/16, 344 1435, 45/10, 350	89.61:95.52 96.97:97.35	92.62 92.75
NewHandPD	320: 320 320: 620	96.09 97.34	94.37 96.28	0.017 0.013	0.027 0.041	0.104 0.062	320, 0/23, 297 600, 20/8, 312	100.00:92.75 96.83:97.60	95.93 96.06
PaHaW	180:180	90.28	86.94	0.036	0.160	0.265	164, 16/19, 161	91.4:89.19	90.14
UCI	75:75 75:125	100.00 97.50	94.67 92.00	0.038 0.062	0.066 0.074	0.029 0.083	75, 0/0, 75 117, 8/0, 75	100.00:100.00 93.75:100.00	100.0 96.77
All Datasets	935:935 935:2405	93.85 96.71	92.94 96.05	0.007 0.007	0.043 0.026	0.191 0.125	859, 76/37, 898 2311, 94/29, 906	91.84:96.07 96.09:96.93	93.99 93.99

The obtained results clearly demonstrate the excellent performances of our approach for PD diagnosis. Even without data augmentation, our model achieved high performances across multiple datasets, suggesting a discernible link between handwriting patterns and the presence of PD. Accuracy on the larger HandPD and NewHandPD datasets reached up to 94.74%, demonstrating the relevance of our approach. While smaller datasets, such as UCI and PaHaW, exhibited higher variance, likely due to their limited size, their performance also reinforces the connection between handwriting and PD. This baseline performance establishes the viability of handwriting as a potential diagnostic marker.

However, the results also highlight the challenges posed by limited data and class imbalance, common issues in medical datasets. While reasonable accuracy was achieved, further analysis of sensitivity and specificity revealed potential biases. Models trained on the original data sometimes struggled to achieve a perfect balance between correctly identifying PD patients (sensitivity) and healthy individuals (specificity). This is where data augmentation becomes a crucial tool. Indeed, our experiments with data augmentation (refer to Table 4), significantly enhance the diagnostic capabilities of the system, acting as a powerful technique to address these challenges. By generating synthetic samples, augmentation effectively increases the size and balance of the training data, allowing the models to learn more robust and generalized patterns. The impact is clearly visible across all datasets.

With augmentation, a substantial improvement in accuracy is observed. On the larger HandPD and NewHandPD datasets, accuracy surpasses 97%, demonstrating the power of this technique to refine diagnostic precision. When all the datasets are combined, the metric reflects this trend, with average accuracy increasing from 89.02% without augmentation to 96.05% with augmentation. This improvement underscores the value of data augmentation for extracting the full diagnostic potential from handwriting data.

Beyond accuracy, data augmentation significantly improves the balance between sensitivity and specificity. By mitigating the effects of class imbalance, augmentation allows the models to more reliably identify both PD patients and healthy individuals. This is crucial for a diagnostic tool, ensuring that the system is both accurate and equitable in its classifications. The impact of augmentation is also reflected in the loss values. Lower loss values in the augmented models indicate better convergence and learning, suggesting that the models are more effectively capturing the underlying patterns associated with PD. Furthermore, the reduced standard deviation of accuracy across multiple runs indicates that augmentation contributes to more stable and reliable performance.

By analyzing the confusion matrices generated for each dataset, we identified specific samples that our model misclassified (see Figure 3). These misclassified samples exhibited high intra-class variability, making it difficult, even for human experts, to accurately assign them to their respective classes. This suggests that the model's behavior is largely consistent with human judgments based on the available data.

Overall, our model demonstrates strong performance and reliability in detecting PD through handwriting analysis. The ability to achieve high accuracy rates, even on challenging datasets with high intra-class variability, highlights the effectiveness of our approach.

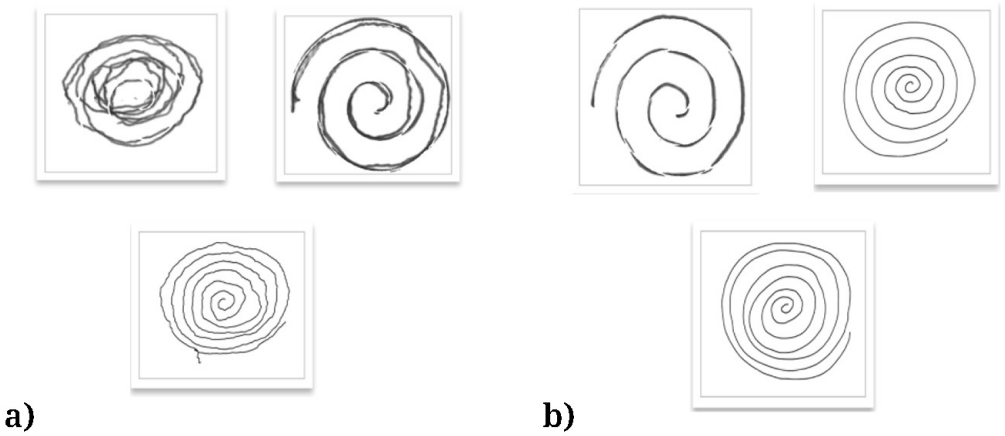

Fig. 3. Our model false acceptance (a) and false rejection (b) samples

5 CONCLUSION

In conclusion, this paper has explored the feasibility of diagnosing PD through offline handwriting analysis, focusing on the motor symptom of hand tremors. By proposing an exploration of offline handwriting analysis for PD diagnosis and presenting a novel deep learning model architecture, this study contributes to the growing body of research in this area. The experiments conducted on four different datasets demonstrate a high accuracy rate in classifying PD and healthy control subjects based on handwriting traces, highlighting the potential of offline handwriting analysis as a non-invasive and cost-effective diagnostic tool for PD. Future research should explore the robustness and generalizability of the approach across diverse populations and refine the model to improve diagnostic accuracy. Overall, this study underscores the importance of leveraging digital health technologies to advance detection and management strategies for Parkinson's disease.

6 REFERENCES

[1] W. Poewe *et al.*, "Parkinson disease," *Nat Rev Dis Primers*, vol. 3, no. 1, p. 17013, 2017. https://doi.org/10.1038/nrdp.2017.13

[2] M. P. Caligiuri, H.-L. Teulings, J. V. Filoteo, D. Song, and J. B. Lohr, "Quantitative measurement of handwriting in the assessment of drug-induced parkinsonism," *Hum. Mov. Sci.*, vol. 25, no. 4, pp. 510–522, 2006. https://doi.org/10.1016/j.humov.2006.02.004

[3] M. Kaur and K. Saini, "Forensic examination of effects of Parkinsonism on various handwriting characteristics," *Science & Justice*, vol. 62, no. 1, pp. 10–20, 2022. https://doi.org/10.1016/j.scijus.2021.10.009

[4] A. Bensefia, C. Djeddi, and A. Hannousse, "Parkinson disease detection via handwriting analysis," in *10th International Conference on Machine Learning Technologies (ICMLT 2025)*, Helsinki, 2025, pp. 510–514. https://doi.org/10.1109/ICMLT65785.2025.11193333

[5] Md. A. Islam, Md. Z. Hasan Majumder, Md. A. Hussein, K. M. Hossain, and Md. S. Miah, "A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets," *Heliyon*, vol. 10, no. 3, p. e25469, 2024. https://doi.org/10.1016/j.heliyon.2024.e25469

[6] S. Mitra, P. Mohan Pandey, V. Pandey, T. Sharma, and R. Nair, "A fine-tuned transfer learning approach for Parkinson's disease detection on new hand PD dataset," in *International Conference on Applied Technologies*, M. Botto-Tobar, M. Zambrano Vizuete, S. Montes León, P. Torres-Carrión, and B. Durakovic, Eds., Cham: Springer Nature Switzerland, 2024, pp. 86–98. https://doi.org/10.1007/978-3-031-58953-9_7

[7] S. Agrawal and S. P. Sahu, "Image-based Parkinson disease detection using deep transfer learning and optimization algorithm," *International Journal of Information Technology*, vol. 16, no. 2, pp. 871–879, 2024. https://doi.org/10.1007/s41870-023-01601-3

[8] N. Wang *et al.*, "A coordinate attention enhanced swin transformer for handwriting recognition of Parkinson's disease," *IET Image Process*, vol. 17, no. 9, pp. 2686–2697, 2023. https://doi.org/10.1049/ipr2.12820

[9] I. Kamran, S. Naz, I. Razzak, and M. Imran, "Handwriting dynamics assessment using deep neural network for early identification of Parkinson's disease," *Future Generation Computer Systems*, vol. 117, pp. 234–244, 2021. https://doi.org/10.1016/j.future.2020.11.020

[10] S. Naz, I. Kamran, S. Gul, F. Hadi, and F. Khalifa, "Multi-model fusion of CNNs for identification of Parkinson's disease using handwritten samples," *IEEE Access*, vol. 11, pp. 135600–135608, 2023. https://doi.org/10.1109/ACCESS.2023.3337804

[11] E. Valla, S. Nõmm, K. Medijainen, P. Taba, and A. Toomela, "Tremor-related feature engineering for machine learning based Parkinson's disease diagnostics," *Biomed Signal Process Control*, vol. 75, p. 103551, 2022. https://doi.org/10.1016/j.bspc.2022.103551

[12] M. Diaz, M. A. Ferrer, D. Impedovo, G. Pirlo, and G. Vessio, "Dynamically enhanced static handwriting representation for Parkinson's disease detection," *Pattern Recognit Lett*, vol. 128, pp. 204–210, 2019. https://doi.org/10.1016/j.patrec.2019.08.018

[13] R. Lamba, T. Gulati, K. A. Al-Dhlan, and A. Jain, "A systematic approach to diagnose Parkinson's disease through kinematic features extracted from handwritten drawings," *J Reliab. Intell. Environ.*, vol. 7, no. 3, pp. 253–262, 2021. https://doi.org/10.1007/s40860-021-00130-9

[14] S. Xu and Z. Pan, "A novel ensemble of random forest for assisting diagnosis of Parkinson's disease on small handwritten dynamics dataset," *Int. J Med. Inform.*, vol. 144, p. 104283, 2020. https://doi.org/10.1016/j.ijmedinf.2020.104283

[15] M. Diaz, M. Moetesum, I. Siddiqi, and G. Vessio, "Sequence-based dynamic handwriting analysis for Parkinson's disease detection with one-dimensional convolutions and BiGRUs," *Expert Syst. Appl.*, vol. 168, p. 114405, 2021. https://doi.org/10.1016/j.eswa.2020.114405

[16] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Decision support framework for parkinson's disease based on novel handwriting markers," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 508–516, 2015. https://doi.org/10.1109/TNSRE.2014.2359997

[17] M. M. Al-Nawashi, O. M. Al-Hazaimeh, and M. Kh. Khazaaleh, "A new approach for breast cancer detection- based machine learning technique," *Applied Computer Science*, vol. 20, no. 1, pp. 1–16, 2024. https://doi.org/10.35784/acs-2024-01

[18] N. Y. Gharaibeh, A. A. Abu-Ein, O. M. Al-hazaimeh, K. M. O. Nahar, W. A. Abu-Ain, and M. M. Al-Nawashi, "Swin transformer-based segmentation and multi-scale feature pyramid fusion module for Alzheimer's disease with machine learning," *Int. J. Online Biomed. Eng.*, vol. 19, no. 4, pp. 22–50, 2023. https://doi.org/10.3991/ijoe.v19i04.37677

[19] O. Al-Hazaimeh, A. Abu-Ein, N. Tahat, M. Al-Smadi, and M. Al-N'awashi, "Combining artificial intelligence and image processing for diagnosing diabetic retinopathy in retinal fundus images," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 13, pp. 131–151, 2022. https://doi.org/10.3991/ijoe.v18i13.33985

[20] F. Abu-Amara, H. Mohammad, and A. Bensefia, "Robot-based therapy for improving academic skills of children with autism," *International Journal of Information Technology*, vol. 16, no. 6, pp. 3371–3380, 2024. https://doi.org/10.1007/s41870-024-01883-1

[21] F. Brescia, A. Santana, M. Diaz, G. Vessio, M. Ferrer, and G. Castellano, "Integrating robotic kinematics and dynamics with online handwriting features for dysgraphia classification," *Biomed. Signal Process Control*, vol. 112, p. 108560, 2025. https://doi.org/10.1016/j.bspc.2025.108560

[22] C. R. Pereira *et al.*, "A new computer vision-based approach to aid the diagnosis of Parkinson's disease," *Comput Methods Programs Biomed*, vol. 136, pp. 79–88, 2016. https://doi.org/10.1016/j.cmpb.2016.08.005

[23] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Sao Paulo, Brazil, 2016, pp. 340–346. https://doi.org/10.1109/SIBGRAPI.2016.054

[24] M. Isenkul, B. Sakar, and O. Kursun, "Improved spiral test using digitized graphics tablet for monitoring Parkinson's Disease," in *2nd International Conference on E-Health and TeleMedicine-ICEHTM*, Istanbul, Turkey, 2014, pp. 171–175. https://doi.org/10.13140/RG.2.1.1898.6005

[25] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," *Artif. Intell. Med.*, vol. 67, pp. 39–46, 2016. https://doi.org/10.1016/j.artmed.2016.01.004

7 AUTHORS

Ameur Bensefia is with the Higher Colleges of Technology, Abu Dhabi, UAE (E-mail: abensefia@hct.ac.ae).

Chawki Djeddi is with the Laboratoire de Vision et d'Intelligence Artificielle (LAVIA), Université Larbi Tebessi, Tébessa, Algeria; Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes (LITIS), Université de Rouen-Normandie, Rouen, France.

Abdelhakim Hannousse is with the Laboratoire des Problèmes Inverses: Modélisation, Information et Systèmes (PI: MIS), Université 8 Mai 1945, Guelma, Algeria.

Moizes Diaz is with the Instituto Universitario para el Desarrollo Tecnológico y la Innovación en Comunicaciones, Universidad de Las Palmas de Gran Canaria, Campus de Tafira, Las Palmas de Gran Canaria, Spain.