Journal of Computer Science

Original Research Paper

Fusing CNN Models for Improved Parkinson's Disease Detection from Handwritten Features

Sabrina Benredjem and Tahar Mekhaznia

Laboratory of Mathematics, Informatics and Systems (LAMIS), Echahid Cheikh Larbi Tebessi University, Tebessa, Algeria

Article history
Received: 01-07-2024
Revised: 27-07-2024
Accepted: 29-07-2024

Corresponding Author:
Sabrina Benredjem
Laboratory of Mathematics, Informatics and Systems (LAMIS), Echahid Cheikh Larbi Tebessi University, Tebessa, Algeria
Email: sabrina.benredjem@univ-tebessa.dz

Abstract: Neurodegenerative Diseases (NGD) are a group of progressive neurological disorders, including cognitive decline as Alzheimer's Disease (AD) and Parkinson's Disease (PD) that result in the progressive loss of neuronal structure or function. This can lead to a variety of symptoms, in this study we are interested in PD one of the neurological diseases that causes the loss of dopamine-producing neurons in the brain, leading to movement disorders., movement disorders, and dementia. In this study, we are interested in PD, one of the neurological diseases that causes the loss of dopamine-producing neurons in the brain, leading to movement disorders. Early diagnosis of PD seems to be the best way to improve the quality of the patient's life by prescribing the appropriate treatment. The relevant observed symptoms are often subtle; these include slow movement, decreased performance in carrying out daily tasks, tremors, muscle stiffness, and various other psychological symptoms. Handwriting or drawing analysis is one of the dominant mechanisms supporting the early diagnosis and assessment of PD. Based on that, to improve the reliability of Parkinson's Disease (PD) detection, we implemented various data augmentation techniques to increase the size of the dataset. We then deployed and trained various architectures of deep Convolutional Neural Networks (CNNs), each capturing different salient features and aspects of the input data due to their unique layout and structure. We then carefully selected promising feature vectors and applied various early fusion strategies before the final classification step. Early fusion combines the feature vectors extracted by multiple CNNs at an early stage, allowing the classification model to learn and recognize different representations of the data provided by these CNNs. This technique is very beneficial as it improves the model's ability to capture a wide range of features and improves overall system performance. Our experimental results demonstrate that the fusion of frozen features from multiple deep CNN models yields a substantial improvement in accuracy, achieving an impressive exactness rate of 96.29%. This performance surpasses that of individual CNN models and even outperforms other state-of-the-art approaches, highlighting the effectiveness of our fusion-based strategy in enhancing PD detection accuracy.

Keywords: Parkinson's Disease, Features Fusion, Deep Learning, Handwriting Analysis, Transfer Learning

Introduction

The nervous system is a complex network of nerve cells responsible for controlling all bodily functions. It comprises the brain, the spinal cord, and the peripheral nervous system. It enables us to interact with ours. Environment, learn and adapt to new situations, and maintain our internal homeostasis. Neurons are specific cells that constitute the basic structural and functional block of the nervous system. Neurons cannot be divided or regenerated; when damaged, they alter the brain activity as paralysis, blindness, or cognitive impairment. Within specific conditions (such as old age), it leads to NGD. These diseases are usually chronic, long-lasting, and incurable. They can affect people of all ages, but they occur commonly in older adults. Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and multiple sclerosis are the most frequent NGD types (Wu *et al*., 2023).

Parkinson's Disease (PD) is a progressive, incurable, chronic neurodegenerative disorder. It affects the movement control part of the central nervous system. It is characterized by motor symptoms that worsen over time due to the presence of Lewy bodies and abnormal protein deposits in the brain. British physician James Parkinson first described PD in his 1817 essay "An Essay on Parkinsonism" (Hussain *et al*., 2018). The PD progresses at different rates among people, leading to various disorder symptoms. It is estimated that currently, PD affects approximately 5 million people worldwide. In 2030, this number is expected to rise to 9 million. It is also noted that more than 1% of people over the age of 60 suffer from Parkinson's disease (Brakedal *et al*., 2022).

There is no definitive test for PD that can accurately diagnose the condition within living individuals (Li *et al*., 2021). However, protein deposits can be definitively identified through the post-mortem examination of brain tissue (Goldma, 2016). This indicates an extremely tardy diagnosis in the disease's pathophysiological development. Therefore, the great challenge facing medical research lies in the necessity of finding ways to detect the disease in the early stage, to slow or even stop its progression in the early stages.

Context and Objectives

Neurologists typically diagnose Parkinson's Disease (PD) through clinical observation of the primary motor symptoms, which include tremors, rigidity, and bradykinesia. These cardinal symptoms, however, usually become apparent only after the disease has significantly progressed, with the loss of about 50-60% of dopaminergic neurons in the substantia nigra before the symptoms are visible (Cheng *et al*., 2010). This progression limits the effectiveness of early intervention strategies, highlighting the urgent need for earlier diagnostic methods to slow disease progression and enhance patient quality of life.

Current diagnostic techniques face several challenges. Clinical observation, while essential, is inherently subjective and varies with the neurologist's expertise, which can lead to inconsistencies in diagnosis (Tolosa *et al*., 2021). Neuroimaging methods like MRI and PET scans offer more objective insights but are costly, less accessible, and more useful in the disease's later stages (Aderinto *et al*., 2023). Moreover, biomarker-based approaches are still in the experimental stages and lack standardized protocols for clinical application.

Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) present promising avenues for improving early detection of PD. Notably, deep learning models have shown potential in identifying subtle changes in medical data such as handwriting and speech, which are affected early in PD's course (Garcia Santa Cruz *et al*., 2023). Handwriting analysis, in particular, has emerged as a non-invasive and cost-effective method for detecting early motor impairments associated with PD, revealing distinct changes such as "Micrographia," characterized by small and choppy handwriting (Senatore *et al*., 2022).

To address the limitations of current diagnostic methods, our study proposes leveraging deep convolutional neural networks (CNNs) combined with advanced machine learning techniques to analyze handwriting features. By employing a fusion-based approach that integrates multiple CNN architectures, we aim to enhance the accuracy of PD detection. This methodology not only promises to improve diagnostic performance but also offers a scalable solution for early screening of PD.

Our research seeks to bridge the gap in early PD diagnosis by utilizing deep learning and handwriting analysis to develop a reliable, non-invasive diagnostic tool. This tool is designed to be easily integrated into clinical practice, offering an innovative approach to detect and monitor the progression of Parkinson's disease.

Literature Review

Recent advances in artificial intelligence technologies have generated significant excitement and a unique buzz in a variety of fields, including medicine. AI-based methodologies are now being widely used to diagnose and predict a wide range of diseases at an early stage. Many studies have been conducted on the use of AI to diagnose PD using various datasets.

Khatamino *et al*. (2018) proposed a CNN-based tool to analyze spiral patterns in PD patients using two datasets. Through cross-validation, the accuracy of PD classification reaches more than 88%.

Rios-Urrego *et al*. (2019) proposed a classification model for detecting PD and healthy subjects using kinematic, geometric, and nonlinear dynamic analysis. The model achieved an accuracy rate of 93.1%, emphasizing the need for further validation.

Moetesum *et al*. (2019) developed a system that uses multiple networks to extract features from different writing movement samples, achieving 83% accuracy.

Chakraborty *et al*. (2020) developed a system using 2D CNN to analyze handwritten patterns (spirals and waves) in PD patients and healthy subjects and achieved a remarkable overall accuracy of 93.3%.

Shaban, (2020) introduced a technique for PD detection using pre-trained VGG 19 models that make use of spiral and wave drawing. For spiral images, their method had accuracy.

Gazda *et al*. (2022) proposed a recently pre-trained neural network approach for detecting PD from offline handwriting. Their innovative architecture achieved an impressive 94.7% accuracy in the classification of PD from offline handwriting.

Discussion

Some models excel at seamlessly incorporating multiple networks and feature sets, emphasizing the importance of feature engineering and network architecture. Furthermore, incorporating pre-trained models such as VGG 19 has shown promising results in partial discharge detection and simplifies the overall model development process. The limitations imposed by previous studies vary and must be taken into account to avoid falling into them and achieve the desired results. The data sets of several studies were relatively small, underscoring the need for further validation with larger and more diversified data sets to guarantee and improve generalizability and flexibility. A second limitation of current approaches to detecting PD handwriting-based methods lies in their limited generalizability to diverse populations and handwriting styles. Another limit is the limited exploration of advanced techniques, such as transfer learning, which represents a missed opportunity for enhancing PD detection accuracy. Without the need for an abundance of data or computing resources, researchers may develop more precise models for PD diagnosis by employing transfer learning.

Proposed Model

PD is presently incurable, but its symptoms can be effectively managed to slow the disease progression. Timely diagnosis proves crucial, not only for addressing motor function fluctuations but also for establishing an optimal medication regimen. Prior studies have significantly advanced Parkinson's disease detection, showcasing, the substantial potential of machine learning and deep learning techniques. These studies have utilized a diverse range of analytical methods, including kinematic, geometric, and Convolutional Neural Network (CNN)-based approaches. Our primary objective is to facilitate the initial diagnosis of Parkinson's disease, aiding doctors in gaining a comprehensive understanding of their patients promptly. We aim to provide a practical and affordable detection and diagnostic tool for Parkinson's disease, enabling the identification of nervous system abnormalities at an early stage. This not only helps in preserving medical expenses but also plays a vital role in maintaining human health. Furthermore, our goal is to make a self-diagnosis test tool available to the general public, based on handwriting samples. Ultimately, our efforts are directed towards enhancing accessibility to early diagnosis, promoting timely intervention, and positively impacting both healthcare and financial aspects for individuals affected by Parkinson's disease. To begin achieving our goal, we have created a structured series of stages that require passing through to attain the targeted goal as shown in Fig. (1)

Used Tool

To overcome the mentioned weaknesses, we propose an appr model that involves a multi-step process for analyzing handwritten images of Parkinsonian and non-Parkinsonian patients. This section provides a thorough description of the tools and techniques used in the development of the proposed PD detection model.

Transfer Learning

Transfer learning is a machine learning technique that applies knowledge learned from a source task to a target task. Source tasks are typically well-defined problems with large amounts of labeled data, while target tasks are related problems with less or no labeled data. The goal of transfer learning is to use knowledge from the source task to improve the performance of the model on the target task. Transfer learning can also save a lot of training time, as new models don't need to be trained from scratch. It allows users to apply a whole new set of data to solve completely different problems. It allows the user to specify the dimensions of the final layer as desired. What's more, transfer learning methods not only allow users to modify the dimensions of the output layer but also enable users to fine-tune other hyperparameters and weights of other layers in the pre-trained model.

VGG16

The VGG16 is a Convolutional Neural Network (CNN) architecture developed by Karen Simonyan and Andrew Zisserman at the University of Oxford. It is one of the more prevalent image feature extraction techniques as it can extract a large amount of data while producing good accuracy. It is a deep network with 16 layers of convolution and pooling operations. The convolutional layer uses a 3×3 filter with a stride of 1, followed by a ReLU activation function. The pooling layer uses a max pooling operation with 2×2 filters and stride 2. The network also has two fully connected layers at the end and the last layer has a softmax activation function. In this study, we freeze training and remove the fully connected layers in the multi-layer convolutional neural network VGG16.

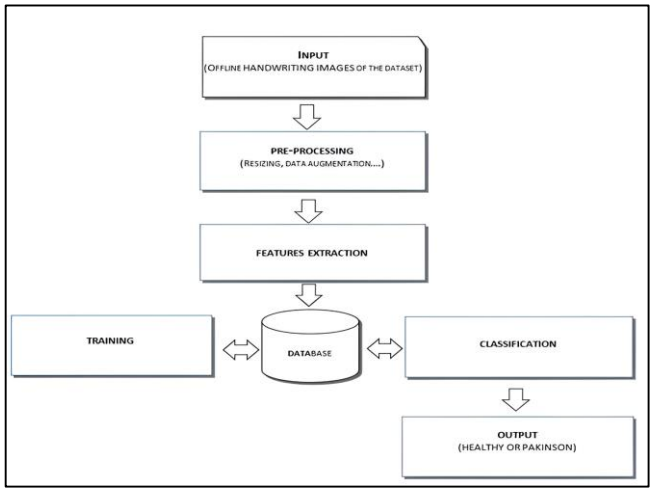

Fig. 1: The workflow of the proposed model

AlexNet

The AlexNet, a deep convolutional neural network with 8 layers, revolutionized computer vision in 2012 by winning the ImageNet image recognition challenge. This architecture, created by Krizhevsky, Sutskever, and Hinton, used clever techniques like ReLU activations, max-pooling, and local response normalization to train on the massive ImageNet dataset. AlexNet's success marked a turning point, demonstrating the power of deep learning for image classification and inspiring future advancements in CNNs and deep learning research as a whole.

ResNet50

The ResNet50 is a 50-layer deep convolutional neural network, famous for introducing residual learning blocks to solve vanishing gradient problems. It excels at image recognition tasks and takes advantage of skip connections and batch normalization to improve training stability and convergence. Typically used with transfer learning, ResNet50's architecture and training methods make it a powerful tool for various computer vision applications.

Random Forest

RF is a classification algorithm built based on several decision trees composed of each of a different subset followed by a selection of leaf nodes. This classifier applies clustering and randomness to generate a forest of uncorrelated trees. The committee outperforms any individual tree in making predictions and the majority vote determines the ranking.

Architecture

The proposed model for PD diagnosis is structured into distinct components to ensure a comprehensive approach. This model involves several key stages: Data gathering, pre-processing, feature extraction, model training, testing, and classification. Initially, data is meticulously collected from 150 participants at the Department of Neurology at the Constantine University Hospital Center Ibn Badis, adhering to a standardized protocol to ensure consistency and reliability. Pre-processing steps are crucial to clean, normalize, and balance the dataset, ensuring data integrity and uniformity across samples. The model employs advanced neural network architectures-AlexNet, VGG16, and ResNet50-for feature extraction from a variety of handwriting images, capturing intricate patterns indicative of Parkinson's disease. Following feature extraction, a fused feature vector is created and inputted into a Random Forest (RF) classifier for the final classification task. The RF classifier aggregates predictions from multiple decision trees to enhance accuracy and reliability. The outcomes are then categorized into two distinct groups: Healthy and Parkinson's, providing a detailed and systematic process for detecting and categorizing health conditions based on the model's analysis. This comprehensive approach, from data collection to classification, ensures a robust model for diagnosing PD, facilitating accurate and reliable clinical applications.

Input: Images (dataset of images).

Output: Prediction of the result:

1- Load your image dataset.
2- For each image in the dataset Read the image in color. Resize the image to 256*256 (To create floating point values that show better results). End for each
3- Convert image lists to arrays
4- Encode the labels into integers
5- Split data into the test and training datasets
6- Binarization of dataset images
7- Load the VGG16, ResNet50, and AlexNet models for additional feature extraction.
8- For each layer in loaded layers: layer. trainable = False (Make layers non-trainable trainable parameters = 0) end for each
9- Extract features from the training data from the convolutional network for the RF classifier.
10- Fuse the extracted features from VGG16, ResNet50, and AlexNet into a single feature vector.
11- Import RF classifier
12- Train the model on the training
13- Prediction using test data

The schematic diagram of the proposed transfer learning for Parkinson's disease detection is shown in Fig. (2).

Dataset

The data was collected following a standardized protocol to ensure consistency and reliability. This protocol included several key steps designed to minimize variability and maximize the quality of the collected data. First, a total of 150 participants were carefully selected based on inclusion and exclusion criteria relevant to the study's goals. These participants were then provided with printed sheets containing a series of predefined tasks that were to be completed during the data collection session.

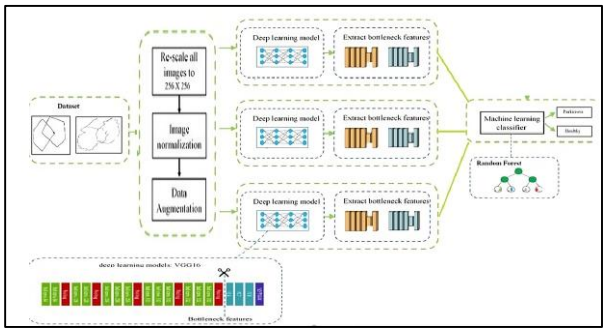

Fig. 2: The schematic diagram of the proposed system

The data acquisition process was conducted at the Department of Neurology at the Constantine University Hospital Center (CHU) Ibn Badis, a facility equipped with the necessary resources and environment for such a study. During the data collection session, participants were allowed to take breaks as needed to prevent fatigue, ensuring that their performance remained consistent. The tasks were repeated to gather sufficient data for reliable analysis.

Throughout the process, we were present to supervise the session, provide assistance, and ensure adherence to the standardized protocol. This approach not only helped in maintaining the quality and reliability of the data but also ensured that any potential issues could be promptly addressed.

All collected data was immediately saved and backed up to secure the information and prevent any loss. This meticulous approach to data collection ensured that the resulting dataset was robust and suitable for subsequent analysis.

Figure (3) illustrates a segment of the standardized protocol, showcasing the handwriting tasks administered to the participants. These tasks included writing a predetermined text and drawing various graphics. To ensure the capture of performance variability, each task was performed twice by each participant. Participants received clear verbal instructions and were allowed to practice the tasks before the official recording session.

The completed sheets were meticulously collected and subsequently digitized using a high-resolution scanner. This process was essential to convert the physical handwriting data into a digital format suitable for advanced analysis. The high resolution of the scanner ensured that even the finest details of the handwriting were preserved, enabling precise and accurate analysis.

Figure (4) provides an overview of the clinical data collection process related to Parkinson's Disease (PD), which was an integral part of the acquisition protocol. Established clinical scales were employed to gather comprehensive information on various dimensions of PD, including disease duration, educational background, hand dominance (laterality), and neuropsychiatric profiles.

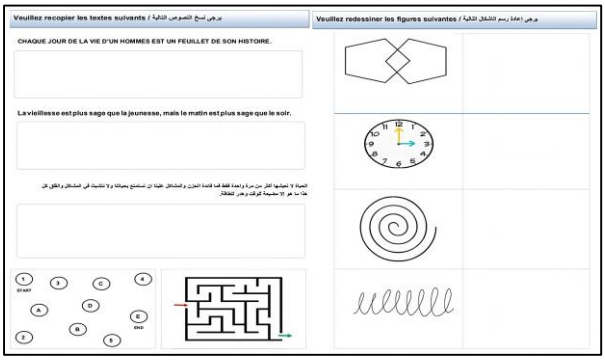

Fig. 3: Data collection form

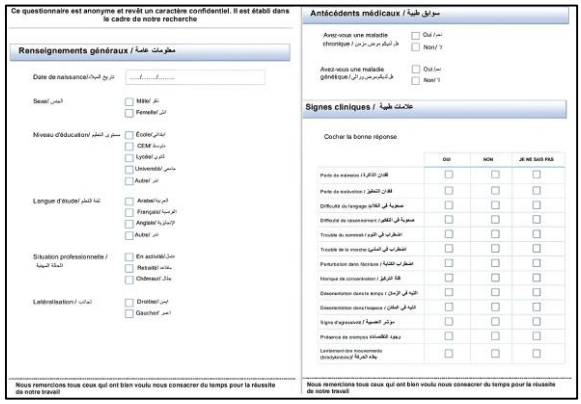

Fig. 4: Clinical data form

Incorporating clinical data into the acquisition protocol significantly enhances the value of the dataset. This approach allows researchers to investigate potential correlations between handwriting performance and clinical characteristics of PD, providing a more holistic understanding of the disease. The enriched dataset facilitates a multifaceted analysis, offering deeper insights into how PD affects handwriting and potentially unveiling new avenues for clinical assessment and intervention.

Data Processing

Pre-processing of the dataset images may be required to improve its quality to extract useful features. Preprocessing involves a variety of techniques, depending on the image's characteristics. In this section, two main methods are used as shown in Fig. (5).

Binarization: The technique converts the filtered image to binary format; it consists of adjusting pixel intensities to 1 and 0 values according to a given threshold. We calculate the threshold value by averaging all pixel intensities in the document.

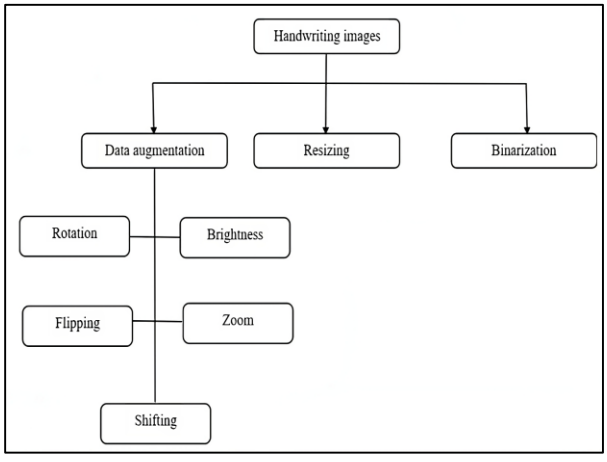

Fig. 5: The image pre-processing methods

The following algorithm shows how to convert an image to a binary value.

Algorithm: Binarization

Input: the image (grayscale)

Output: the image (black and white)
pixel_intensity_sum = 0

pixel_count = 0

For each pixel image[I][J] in the images do

Pixel_Intensity_Sum=Pixel_Intensity_Sum+ image[I][J]
Pixel_Count=Pixel_Count+1 End Foreach

Average_Intensity= Pixel_Intensity_Sum/ Pixel_Count For each pixel image[I][J] in the

images do

If (image[I][J] >=Average_Intensity) image[I][J] =255 (WHITE)

Else

image[I][J] =0 (BLACK)

In Fig. (6), image (A) offers a preview of a sample of images from the dataset, showing original visual details. In contrast, image(B) reflects the same sample after the full application of the binarization algorithm. In the image the (A), the outlines of the characters are blurred. This is because the image contains pixels with different brightness values. Applying the binarization algorithm clarifies the character outlines. This is because the outline pixels are generally lighter or darker than the background pixels.

Data augmentation: We implemented a data augmentation strategy to significantly increase the diversity of PD handwritten samples within the dataset. Dataset. This strategy is one of the best practices to reduce overfitting in the research community. In this study, we geometrically translated the images to increase the dataset size. Data augmentation included operations such as rotation, zooming, flipping, and shifting (both width and height shifts). Table (1) summarizes the data augmentation parameters and their corresponding values used in this study.

Figure (7) illustrates the data augmentation process on a sample of data drawing "pentagon.

Table (2) displays the number of images in each category of drawing data before and after applying data augmentation techniques. The difference between columns reflects the crucial role data augmentation techniques play in determining the model's effectiveness.

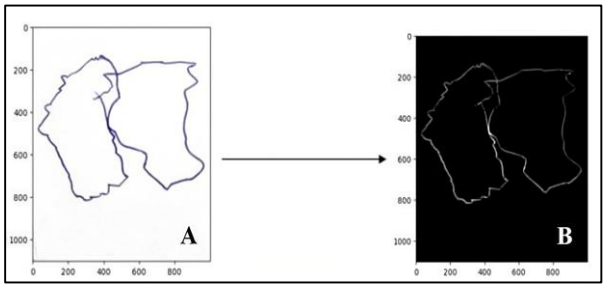

Fig. 6: Example of binarization

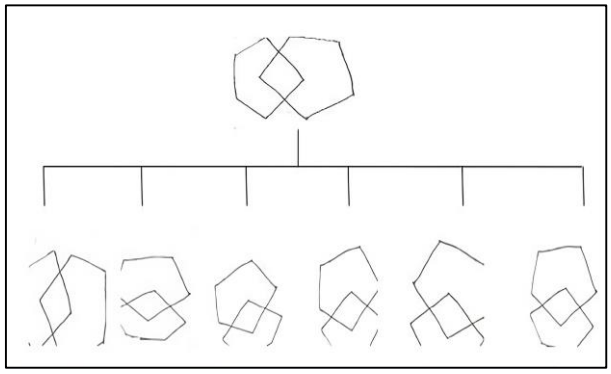

Fig. 7: Example of augmented image

Table 1: Data augmentation parameters

Augmentation parameters	Setting
Width shift range	0.1
Height shift range	0.1
Rotation range	360
Horizontal flip	True
Vertical flip	True
Zoom rang	0.2

Table 2: Number of original and augmented images

Drawing	Before	After
Spiral	156	590
Pentagon	150	440
Clock	132	320
" L " Series	156	432

Feature Extraction and Fusion

Feature extraction is a pivotal process in machine learning, dedicated to the selection and extraction of pertinent information or features from raw data. The objective is to transform the original data into a more informative, concise, and meaningful representation, facilitating better performance in downstream tasks. In this study, we aimed to extract deep and diverse features by leveraging three pre-trained CNN architectures through the frozen feature method. Specifically, we utilized AlexNet, VGG16, and ResNet50 models to generate their respective feature vectors, denoted as FV1-FV3. The frozen feature method involves utilizing pre-trained weights for certain layers without additional training, focusing on predefined layers within each architecture to capture relevant features effectively.

The VGG16 model's structure for extracting features from input images is illustrated in Fig. (8). VGG16 is renowned for its deep architecture consisting of 16 layers, which includes 13 convolutional layers followed by 3 fully connected layers. Each convolutional layer applies a series of filters to the input image, capturing various levels of abstraction, from edges and textures to complex patterns and objects. By utilizing the frozen feature method, we extract feature vectors from these layers without modifying the pre-trained weights, ensuring that the extracted features retain the rich information learned from extensive training on large datasets.

Two options are proposed to combine extracted feature vectors into a single vector using various mathematical operations.

The first fusion scheme involves summing the individual feature vectors extracted from AlexNet (*FVd*1), *VGG*16 (*FVd*2), and ResNet50 (*FVd*3). This approach aggregates the features, capturing a broad spectrum of information from each model. The summation operation is mathematically represented as:

$$FV\ sum\ =\ f_{sum}\ (\ FV\ d1,\ FV\ d2,\ FV\ d3) \quad (1)$$

This method leverages the complementary strengths of each model, potentially enhancing the overall feature representation by combining their diverse perspectives.

In the second fusion scheme, we construct a single comprehensive feature vector by multiplying the individual feature vectors. This approach emphasizes the interactions between features from different models, capturing more nuanced and combined patterns. The product operation is mathematically represented as:

$$FV\ prod\ =\ f_{prod}\ (\ FV\ d1,\ FV\ d2,\ FV\ d3) \quad (2)$$

This method can highlight features that are consistently important across all models, providing a robust feature representation.

Classification

The classification consists of the assignment of each labeled data to a certain category. The model then makes predictions on fresh, unforeseen data after learning the patterns and attributes that distinguish each category. Applications for classification algorithms include fraud detection, spam filtering, sentiment analysis, and image recognition. To improve the results of our model, we used the random forest model as a classifier. We have chosen RF due to its robustness, accuracy, flexibility, scalability, interpretability, and solving the overfitting problem (Impedovo and Pirlo, 2019). It is a highly effective tool for making the precise predictions needed in strategic decisions as shown in Fig. (9).

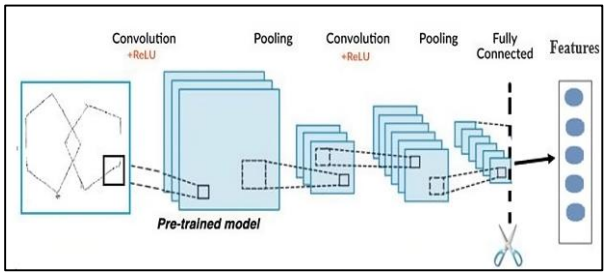

Fig. 8: The feature extraction process

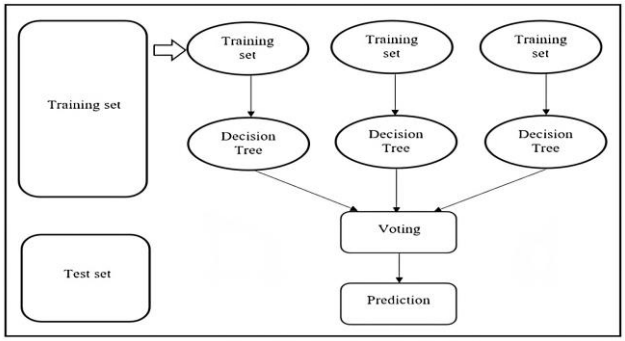

Fig. 9: Random forest classifier

Train and Evaluate the Model

In this section, the image dataset is split into training and test sets, maintaining an 80/20 ratio. Throughout the training process, batches systematically distribute the input data, allowing for seamless execution control by ensuring the data fits into memory efficiently. The provided code snippet 4/4 [==============================]-4s 898ms/step indicates that the training process has completed 4 out of 4 epochs, with each epoch taking approximately 4 sec and 898 m sec per step. This message suggests that the training process has been successfully completed and the model has been trained using the provided training data. To evaluate the performance of the proposed model, we use a set of metrics that provide insight into the predictive capabilities of the model, namely accuracy, recall, precision, and the area under the curve. The following descriptions elaborate on each of these metrics and how they are utilized to evaluate the model's performance.

Accuracy is a measure used to determine the percentage of correct predictions made by a model. A high accuracy score indicates that the model makes globally correct predictions:

$$ACC\ =\ \frac{TP\ +\ TN}{TP\ +\ TN\ +\ FP\ +\ FN}$$

Table (3) summarizes the interpretations of the four terms.

Precision is a metric that measures the proportion of true positive predictions among all positive predictions made by the model. A high precision score indicates that the model is making accurate positive predictions:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall is a metric that measures the proportion of true positive predictions among all actual positive cases in the data. A high recall score indicates that the model is correctly identifying positive cases in the data:

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

The Area Under the Curve (AUC in short) is a metric that quantifies the model's ability to distinguish between positive and negative classes. It is a widely used measure for evaluating the performance of binary classification models.

Performance Results of the Proposed Model

In this study, we conducted an experimental setup and analysis using an ensemble and fusion of feature vectors from multiple CNN architectures. To evaluate the effectiveness of our proposed system, we employed various image pre-processing techniques and data augmentation methods, integrating Convolutional Neural Network (CNN) algorithms such as AlexNet and ResNet50. The experimentation was conducted in Python 3.11 using Jupyter Notebook, with key tools like Keras, OpenCV, and Matplotlib facilitating a seamless workflow. Our computational infrastructure, featuring an Intel (R) Core (TM) i7-7700HQ processor with 16.0 GB of RAM, provided the necessary computational power for efficient model training and testing. We utilized the AlexNet, VGG 16, and ResNet50 architectures, renowned for their robust feature extraction capabilities, to complement each other in extracting rich and diverse features from the dataset. This fusion of feature extraction methods, detailed extensively in the Jupyter Notebook, establishes a comprehensive framework for image classification tasks, highlighting the synergistic combination of deep learning and traditional machine learning techniques.

Table 3: Confusion matrix terminology for binary classification

Term	Interpretation
TP (True Positive)	You predicted positive and it's true
FP (False Positive)	You predicted positive and it's false
FN (False Negative)	You predicted negatively and it's false
TN (True Negative)	You predicted negatively and it's true

Table 4: Performance metrics evaluation of the model

Acc % recall	Precision	AUC
Our Dataset 96,29 0,95	1	0,93
Kaggle Dataset 900,03	0,88	0,90

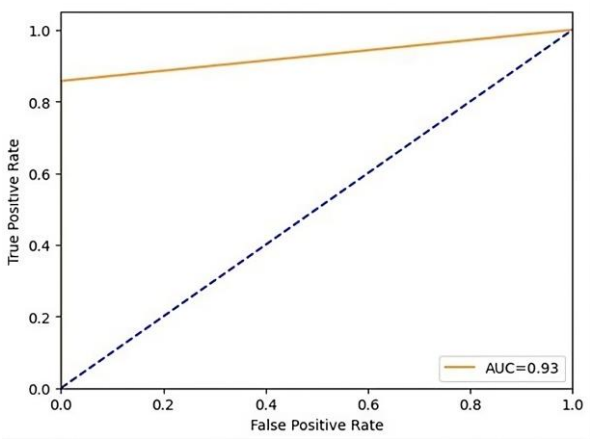

Fig. 10: AUC plot for dataset 1

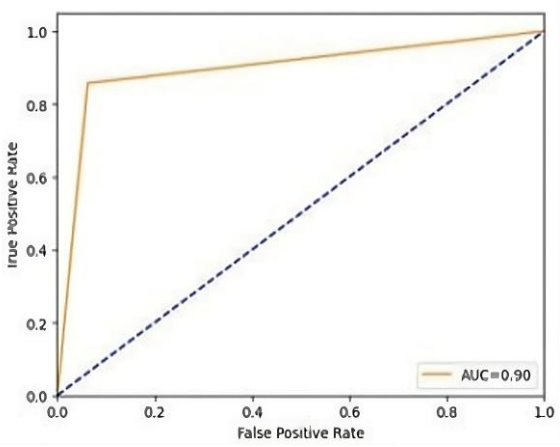

Fig. 11: AUC plot for dataset 2

To evaluate the performance of our model, we extracted diverse performance metrics from the dataset images, as discussed in the previous section. Analyzing these metrics is crucial for understanding the model's performance and its ability to accurately identify and classify patterns in the dataset images, as demonstrated in Table (4).

The ROC curve presented in Fig. (10) further visualizes the model's performance, indicating its high accuracy in classifying instances.

Figure (11) shows the ROC curve for dataset 2, providing a visual representation of the model's performance.

Results and Discussion

To thoroughly assess the performance of our proposed model, it is crucial to compare our results with other PD detection methods that have been previously documented

in the literature. To facilitate this comparison, we have prepared Table (3), which outlines key performance metrics such as accuracy, recall, and precision for a range of networks, including CNN, RF, SVM, AlexNet, and our model. This table offers a more comprehensive insight into the advantages and disadvantages of different approaches, serving as a foundation for evaluating the efficacy of our system.

In Table (5), we present a comprehensive overview of the diagnostic accuracy achieved by our proposed model for detecting PD, as well as several conventional models. The table reveals that existing models demonstrate accuracies ranging from 88-94.70%. However, our model stands in stark contrast, surpassing all others with an impressive accuracy of 96.29% for the addition case and 92.98% for the multiplication case. This substantial performance improvement solidifies the superiority of our model.

These results highlight the superior performance and effectiveness of our proposed model in diagnosing Parkinson's disease, underscoring its potential for advancing clinical applications and medical research. The enhanced accuracy underscores the model's capability to provide reliable and precise diagnostic support, marking a substantial step forward in the field of neurodegenerative disease detection.

To provide a comprehensive comparison of model performance, we have included a detailed graphical representation in Fig. (12). This figure visually illustrates the accuracy metrics of various models, highlighting the superior performance of our proposed model.

As depicted in the graph, the proposed model achieves an impressive accuracy of 96.2%, outperforming all other compared models. The recall and precision metrics further substantiate the robustness and reliability of our model. Specifically, the recall value of 0.95 indicates the model's high sensitivity in correctly identifying positive cases, while the precision value of 1 demonstrates its accuracy in minimizing false positives.

Figure (12) provides a clear and concise visual representation of these comparative results. The bar chart delineates the accuracy percentages of the models, with our proposed model prominently leading the field. This graphical depiction not only underscores the superior accuracy of our model but also highlights its potential as a reliable tool for PD diagnosis.

Table 5: Comparison of performances

Model	Accuracy %	Recall	Precision
Khatamino *et al*. (2018)	88.0	**	**
Rios-Urrego *et al*. (2019)	93.1	**	**
Cascarano *et al*. (2019)	90.0	**	**
Moetesum *et al*. (2019)	83ss	**	**
Chakraborty *et al*. (2020)	93.3	0.94	0.93
Xu and Pan (2020)	89.4	0.84	0.93
Folador *et al*. (2021)	89.6	0.85	0.80
Gazda *et al*. (2022)	94.7	**	**
Proposed model	96.2	0.95	1

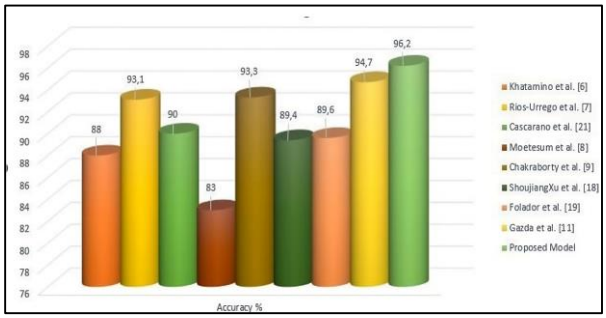

Fig. 12: Comparison of accuracy from proposed and another model

The incorporation of advanced neural network architectures, such as AlexNet, VGG16, and ResNet50, for feature extraction, combined with the RF classifier, contributes significantly to the model's high performance. These components work synergistically to enhance the model's predictive capabilities, ensuring a comprehensive and accurate classification of health conditions.

This research examined the efficacy of ensemble methods in the classification. By utilizing multiple vectors of features derived from different models, the ensemble approach had a significant advantage over a single model that utilized static vectors of features. However, the research has flaws. First, the evaluation utilized only RF for classification. While RFs are impressive, exploring and comparing additional machine learning classifiers will provide a more comprehensive understanding. Second, the train-test split method is common, but it could benefit from the incorporation of K-fold validation. This would enhance the credibility and generalizability of the performance evaluation for the proposed collective model.

Conclusion

PD is characterized by the gradual destruction of nerve cells in the central nervous system, particularly affecting dopamine-producing neurons, leading to various movement disorders. While no medications can halt PD progression in advanced stages, early detection can significantly slow its progression with appropriate treatments. This study proposes a novel model for the early detection of PD using handwriting analysis, leveraging transfer learning techniques with VGG16, AlexNet, and ResNet50 architectures to extract intricate features from drawing images. These features are then classified using the RF algorithm, achieving a high accuracy of 96% in distinguishing between control subjects and PD patients. The model's superior performance compared to existing methods highlights its potential as a valuable tool for early diagnosis, capturing subtle handwriting nuances indicative of PD.

Despite the promising results, several limitations provide opportunities for further research: The current study involved 150 participants from a single location and the reliance on high-resolution scanning and advanced neural network architectures may limit the model's accessibility and offer opportunities for future research.

Future studies should include larger, more diverse populations, optimize the model for accessibility, and conduct multi-center data collection to ensure generalizability. Further exploration of additional discriminative parameters, incorporating metadata like intellectual level and age, and expanding research to other NGDs are also recommended.

In conclusion, our proposed model effectively addresses the study's aims by providing a reliable, non-invasive, and accurate tool for diagnosing PD. By identifying and addressing the study's limitations, future research can further enhance the model's applicability and effectiveness, ultimately contributing to better clinical outcomes and the early detection of PD and potentially other NGDs.

Acknowledgment

The authors express their heartfelt appreciation to the editors for their dedicated efforts in managing the manuscript and to the reviewers for their insightful and constructive comments, which have greatly contributed to enhancing the original submission.

Funding Information

The authors have not received any financial support or funding to report.

Author's Contributions

Sabrina Benredjem: Conceptualization, methodology, data curation, written original draft preparation.

Tahar Mekhaznia: Supervision, visualization, written-reviewed and finalization.

Ethics

The authors confirm that this manuscript has not been published elsewhere and that no ethical issues are involved.

References

Aderinto, N., Olatunji, D., Abdulbasit, M., & Edun, M. (2023). The Essential Role of Neuroimaging in Diagnosing and Managing Cerebrovascular Disease in Africa: A Review. *Annals of Medicine*, *55*(2), 2251490. https://doi.org/10.1080/07853890.2023.2251490

Brakedal, B., Toker, L., Haugarvoll, K., & Tzoulis, C. (2022). A Nationwide Study of the Incidence, Prevalence, and Mortality of Parkinson's Disease in the Norwegian Population. *Npj Parkinson's Disease*, *8*(1), 19. https://doi.org/10.1038/s41531-022-00280-4

Cascarano, G. D., Loconsole, C., Brunetti, A., Lattarulo, A., Buongiorno, D., Losavio, G., Sciascio, E. D., & Bevilacqua, V. (2019). Biometric handwriting analysis to support Parkinson's Disease assessment and grading. *BMC Medical Informatics and Decision Making*, *19*(Suppl 9), 252. https://doi.org/10.1186/s12911-019-0989-3

Chakraborty, S., Aich, S., Jong-Seong-Sim, Han, E., Park, J., & Kim, H.-C. (2020). Parkinson's Disease Detection from Spiral and Wave Drawings using Convolutional Neural Networks: A Multistage Classifier Approach. *2020 22nd International Conference on Advanced Communication Technology (ICACT)*, 298–303. https://doi.org/10.23919/icact48636.2020.9061497

Cheng, H., Ulane, C. M., & Burke, R. E. (2010). Clinical Progression in Parkinson Disease and the Neurobiology of Axons. *Annals of Neurology*, *67*(6), 715–725. https://doi.org/10.1002/ana.21995

Folador, J. P., Santos, M. C. S., Luiz, L. M. D., de Souza, L. A. P. S., Vieira, M. F., Pereira, A. A., & de Oliveira Andrade, A. (2021). On the use of Histograms of Oriented Gradients for Tremor Detection from Sinusoidal and Spiral Handwritten Drawings of People with Parkinson's Disease. *Medical and Biological Engineering and Computing*, *59*(1), 195–214. https://doi.org/10.1007/s11517-020-02303-9

Garcia Santa Cruz, B., Husch, A., & Hertel, F. (2023). Machine Learning Models for Diagnosis and Prognosis of Parkinson's Disease Using Brain Imaging: General Overview, Main Challenges and Future Directions. *Frontiers in Aging Neuroscience*, *15*, 1216163. https://doi.org/10.3389/fnagi.2023.1216163

Gazda, M., Hires, M., & Drotar, P. (2022). Multiple-Fine-Tuned Convolutional Neural Networks for Parkinson's Disease Diagnosis from Offline Handwriting. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, *52*(1), 78–89. https://doi.org/10.1109/tsmc.2020.3048892

Goldma, J. G. (2016). Neuropsychiatric Issues in Parkinson Disease. *CONTINUUM: Lifelong Learning in Neurology*, *22*(4), 1086–1103. https://doi.org/10.1212/CON.0000000000000353

Hussain, R., Zubair, H., Pursell, S., & Shahab, M. (2018). Neurodegenerative Diseases: Regenerative Mechanisms and Novel Therapeutic Approaches. *Brain Sciences*, *8*(9), 177. https://doi.org/10.3390/brainsci8090177

Impedovo, D., & Pirlo, G. (2019). Dynamic Handwriting Analysis for the Assessment of Neurodegenerative Diseases: A Pattern Recognition Perspective. *IEEE Reviews in Biomedical Engineering*, *12*, 209–220. https://doi.org/10.1109/rbme.2018.2840679

Khatamino, P., Canturk, I., & Ozyilmaz, L. (2018). A Deep Learning-CNN Based System for Medical Diagnosis: An Application on Parkinson's Disease Handwriting Drawings. *2018 6th International Conference on Control Engineering and Information Technology (CEIT)*, 1–6. https://doi.org/10.1109/ceit.2018.8751879

Li, H., Pun, C.-M., Xu, F., Pan, L., Zong, R., Gao, H., & Lu, H. (2021). A Hybrid Feature Selection Algorithm Based on a Discrete Artificial Bee Colony for Parkinson's Diagnosis. *ACM Transactions on Internet Technology*, *21*(3), 1–22. https://doi.org/10.1145/3397161

Moetesum, M., Siddiqi, I., Vincent, N., & Cloppet, F. (2019). Assessing Visual Attributes of Handwriting for Prediction of Neurological Disorders—A Case Study on Parkinson's Disease. *Pattern Recognition Letters*, *121*, 19–27. https://doi.org/10.1016/j.patrec.2018.04.008

Rios-Urrego, C. D., Vásquez-Correa, J. C., Vargas-Bonilla, J. F., Nöth, E., Lopera, F. J., & Rafael Orozco-Arroyave, J. (2019). Analysis and Evaluation of Handwriting in Patients with Parkinson's Disease Using Kinematic, Geometrical and Non-Linear Features. *Computer Methods and Programs in Biomedicine*, *173*, 43–52. https://doi.org/10.1016/j.cmpb.2019.03.005

Senatore, R., Marcelli, A., De Micco, R., Tessitore, A., & Teulings, H.-L. (2022). Distinctive Handwriting Signs in Early Parkinson's Disease. *Applied Sciences*, *12*(23), 12338. https://doi.org/10.3390/app122312338

Shaban, M. (2020). Deep Convolutional Neural Network for Parkinson's Disease Based Handwriting Screening. *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*, 1–4. https://doi.org/10.1109/isbiworkshops50223.2020.9153407

Tolosa, E., Garrido, A., Scholz, S. W., & Poewe, W. (2021). Challenges in the Diagnosis of Parkinson's Disease. *The Lancet Neurology*, *20*(5), 385–397. https://doi.org/10.1016/S1474-4422(21)00030-2

Wu, K., Wang, W., Cheng, Q., Li, H., Yan, W., Zhou, F., & Zhang, R. (2023). Pyroptosis in Neurodegenerative Diseases: From Bench to Bedside. *Cell Biology and Toxicology*, *39*(6), 2467–2499. https://doi.org/10.1007/s10565-023-09820-x

Xu, S., & Pan, Z. (2020). A Novel Ensemble of Random Forest for Assisting Diagnosis of Parkinson's Disease on Small Handwritten Dynamics Dataset. *International Journal of Medical Informatics*, *144*, 104283. https://doi.org/10.1016/j.ijmedinf.2020.104283