

# DIAGNOSIS OF PARKINSON'S DISEASE USING BILINGUAL TAMIL AND ENGLISH HANDWRITING BASED TECHNIQUES

Gayathri. M<sup>1</sup>, Hariharan. S. I<sup>2</sup>, Hema. M<sup>3</sup>, Udhayakumar. G<sup>4</sup>

<sup>1,2,3,4</sup>Department of Electronics and Communication Engineering SRM Valliammai Engineering College, SRM Nagar, Kattankulathur, Chengalpattu 603 203, India.

mgayathri2504@gmail.com, hemamuruganantham25@gmail.com

## ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects motor functions, including handwriting. Handwriting analysis has emerged as a promising non-invasive tool for early diagnosis and monitoring of PD. This study explores the use of bilingual handwriting analysis, focusing on Tamil and English scripts, to identify motor impairments associated with PD. Tamil, a curvilinear script, and English, a combination of curves and straight lines, offer complementary insights into the impact of PD on fine motor control. Handwriting samples were collected from participants, including PD patients and healthy controls, in both Tamil and English. Machine learning algorithms, including Support Vector Machines (SVM) and neural networks, were employed to classify handwriting samples as PD or non-PD. The proposed bilingual handwriting-based diagnostic approach offers a cost-effective and non-invasive solution for early detection and monitoring of PD, particularly in multilingual regions.

**Key features:** Stroke dynamics, pressure variation, writing speed, character size

## 1. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disorder that affects millions of people worldwide, causing motor symptoms such as tremors, rigidity, and bradykinesia. Early diagnosis and treatment of PD are crucial to slow down disease progression and improve the quality of life for patients. However, traditional diagnostic methods, such as clinical evaluations and imaging tests, can be time-consuming, expensive, and sometimes inconclusive.

Recent studies have shown that handwriting analysis can be a valuable tool for diagnosing and monitoring neurodegenerative diseases, including PD. Handwriting is a complex motor activity that involves cognitive, motor, and sensory processes, making it an attractive biomarker for PD. Moreover, handwriting analysis is non-invasive, low-cost, and can be performed in a clinical or home setting.

In this study, we propose a novel approach for diagnosing PD using bilingual Tamil and English handwriting-based techniques. Tamil is a widely spoken language in India and other parts of the world, and English is a widely used language globally. By analyzing handwriting patterns in both languages, we aim to develop a more accurate and culturally sensitive diagnostic tool for PD.

The proposed methodology leverages image processing, feature extraction, and machine learning models to detect anomalies in bilingual handwriting. By comparing writing samples from PD patients and healthy individuals, the study seeks to identify language-specific and cross-linguistic markers contribute to the development of non-invasive, cost-effective diagnostic tools tailored to multilingual populations.

## 2. LITERATURE SURVEY

The author in [1] collects the samples of mild to severe parkinson's disease patients along with the samples of healthy person and features of images are extracted. Utilizing corner detection method to judge the fluctuation of trajectory

Research in [2] uses spiral drawings of Parkinson's disease public datasets are used for the investigation and diagnosis. Here 204 spiral drawings along with wave drawings are used for comparative analysis. This paper uses deep learning models, namely DenseNet201, VGG16 to detect the disease.

This paper [3] explains about a novel approach of combining both voice and handwriting analysis. It uses both machine learning and deep learning techniques to enhance diagnostic accuracy. Preliminary findings suggest that the combined analysis of voice and handwriting data using ML and DL techniques yields higher accuracy compared to single-modality approaches.

This work [4] histogram of oriented gradient is combined with Convolutional Neural Network (CNN) to automatically detect Parkinson's disease based on handwriting patterns. The model achieved an accuracy of 87% and an F1 score of 83.21%.

Research in [5] evaluates nine models where those are SVM, Decision Tree, Random Forest, AdaBoost, Gradient Boosting, XGBoost, KNN, CNN, and Tabular Transformer, achieving improved accuracy across all models compared to previous studies, marking a notable advancement in Parkinson's disease classification performance.

The study [6] uses two deep learning models InceptionV3 and Xception to predict the risk of Parkinson's disease based on the hand writing pattern of the drawing. This model is trained on models consisting 4500 images. This tool achieves accuracy rate of 97% which improves the lives of countless people battling this condition.

This paper [7] presents a comprehensive approach for Parkinson disease detection through the integration of voice and handwriting analysis. Uses Convolutional Neural Networks for handwriting detection and Principal Component Analysis for voice analysis, the proposed method aims to accurately identify individuals with Parkinson disease.

This work [8] a cosine annealing scheduler and deep transfer learning are used as a method for handwriting based prediction. It uses samples from both with and without Parkinson's disease, found in NIATS collection. With the use of spiral sketching and CNN, this research hopes to provide a PD detection method.

Review in [9] reports the testing of NeuroDiag, a software-based medical device, for the automated detection of PD using handwriting patterns. NeuroDiag is designed to direct the user to perform six drawing and writing tasks, and the recordings are then uploaded onto a server for analysis. Kinematic information and pen pressure of handwriting are extracted and used as baseline parameters.

This study [10] proposes a hybrid fusion approach that makes use of the visual information derived from different handwriting samples and handwriting templates enhancing the performance in diagnosis of Parkinson's disease.

This project mainly focuses on early diagnosis of Parkinson's disease using different handwriting samples from both the affected and healthy people. It is detected using the machine learning method mainly using the support vector machine (SVM) in python along with many library files.

### 3. PROPOSED SYSTEM

The process begins with the acquisition of a Datasets of Parkinson's Disease, Implying a collection of medical images relevant to the condition. These images then undergo a crucial Data Pre-processing stage, which is vital for enhancing image quality and preparing them for effective model training. This stage involves three key steps: Histogram Equalization, which improves the contrast and visibility of details in the images; Resizing, which standardizes the dimensions of all images to ensure uniformity and compatibility with the model; and Augmentation by Rotation, a technique to artificially expand the dataset by rotating existing images, thereby increasing the model's robustness and reducing the risk of overfitting. The outcome of this pre-processing is a set of Pre-processed Images that are now optimized for analysis.

Subsequently, the process branches into three distinct modelling approaches. The first utilizes Pre-trained Models, leveraging existing deep learning architectures that have been trained on large, general datasets and are fine-tuned for the specific task of Parkinson's Disease detection. The second approach involves building Deep Learning Models from scratch, allowing for customization and optimization of the network architecture to suit the unique characteristics of the medical image data. The third approach employs Machine Learning Models (Using Python), indicating the use of traditional machine learning algorithms implemented with Python libraries. All three model types then undergo a Training phase, where they learn to identify patterns and features indicative of Parkinson's Disease from the pre-processed images. Following training, the models are subjected to Testing and Evaluation using a separate dataset to assess their performance and generalization capabilities. The results from each model are then compared in a Model Comparison stage to determine the most effective approach. Finally, the model demonstrating the highest accuracy and reliability is selected as the Final Model for Parkinson's Disease detection, ready for potential deployment in clinical or research settings.

#### 3.1 FLOW DIAGRAM

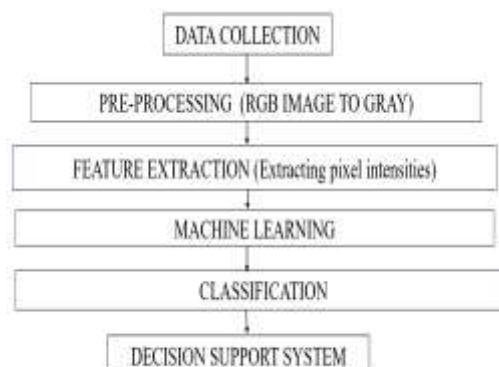


FIG 1. FLOW DIAGRAM

This flow diagram contains mainly of six stages from data collection to decision support system. In Data collection all the input sample images are collected and then they are sent to the pre-processing module where the RGB images are converted into gray image. Then they are passed to feature extraction where it extracts both static and dynamic features from both Tamil and English handwriting samples.

The machine learning module begins by splitting the dataset into training and testing sets. It utilizes support vector machine (SVM) and it uses many library files as. Classification module is the final step in diagnosis of Parkinson's disease. Finally, it reaches decision support system (DSS) which uses the classification results to provide a probability score for PD diagnosis.

### 3.2 BLOCK DIAGRAM

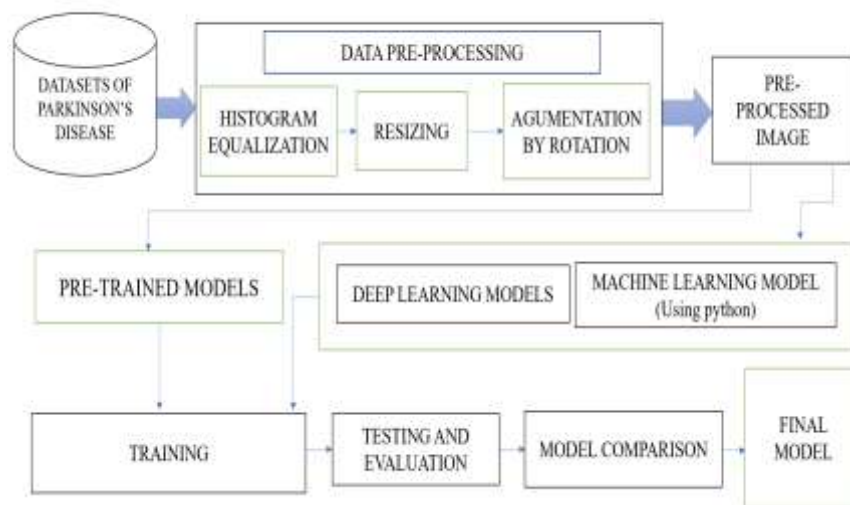


FIG 2. BLOCK DIAGRAM

The process begins with the acquisition of a Datasets of Parkinson's Disease, Implying a collection of medical images relevant to the condition. These images then undergo a crucial Data

Pre-processing stage, which is vital for enhancing image quality and preparing them for effective model training. This stage involves three key steps: Histogram Equalization, which improves the contrast and visibility of details in the images; Resizing, which standardizes the dimensions of all images to ensure uniformity and compatibility with the model; and Augmentation by Rotation, a technique to artificially expand the dataset by rotating existing images, thereby increasing the model's robustness and reducing the risk of overfitting.

Subsequently, the process branches into three distinct modelling approaches. The first utilizes Pre-trained Models, leveraging existing deep learning architectures that have been trained on large, general datasets and are fine-tuned for the specific task of Parkinson's Disease detection.

The second approach involves building Deep Learning Models from scratch, allowing for customization and optimization of the network architecture to suit the unique characteristics of the medical image data.

The third approach employs Machine Learning Models (Using Python), indicating the use of traditional machine learning algorithms implemented with Python libraries. All three model types then undergo a Training phase.

Following training, the models are subjected to Testing and Evaluation using a separate dataset to assess their performance and generalization capabilities.

### 3.3 METHODOLOGY

The proposed system is a bilingual handwriting-based diagnostic tool for Parkinson's disease. The system consists of the following components:

- Data Collection Module
- Preprocessing Module
- Feature Extraction Module
- Machine Learning Module
- Classification Module
- Decision Support System Module

### 3.3.1 DATA COLLECTION MODULE

Collect handwriting samples in both Tamil and English from participants, including healthy individuals and those diagnosed with Parkinson's disease (PD). Participants perform predefined writing exercises such as copying sentences, free handwriting, and drawing spirals. Digital tablets or smart pens capture real-time handwriting dynamics, including stroke pressure, speed, and tremor patterns. For traditional paper-based samples, scanned images are processed for further analysis.

### 3.3.2 PREPROCESSING MODULE

This module enhances handwriting data quality for accurate analysis. It begins with noise reduction and binarization to remove distortions from scanned or digital samples. Normalization ensures uniformity in size and orientation across Tamil and English handwriting. Segmentation techniques separate individual characters or words, enabling precise feature extraction. Additionally, edge detection and contour analysis refine stroke details, while smoothing algorithms reduce unwanted variations. These preprocessing steps prepare the handwriting data for feature extraction and classification, improving the accuracy of Parkinson's disease diagnosis.

### 3.3.3 FEATURE EXTRACTION MODULE

It mainly focuses on identifying key handwriting characteristics that distinguish Parkinson's disease (PD) patients from healthy individuals. It extracts both static and dynamic features from Tamil and English handwriting samples. Static features include letter size, spacing, stroke width, slant, and loop formation, which help analyse script-specific variations. Dynamic features such as writing speed, pressure variation, acceleration, and tremor intensity are captured using digital tablets or smart pens.

### 3.3.4 MACHINE LEARNING MODULE

It is responsible for analysing the extracted handwriting features to classify individuals as Parkinson's disease (PD) patients or healthy controls. This module begins by splitting the dataset into training and testing sets to ensure effective model learning. It uses Support Vector Machine (SVM). The extracted static and dynamic handwriting features serve as input to these models, which learn patterns associated with Parkinson's symptoms such as tremors, inconsistent stroke pressure, and abnormal speed variations.

### 3.3.5 CLASSIFICATION MODULE

The Classification Module is the final stage in diagnosing Parkinson's disease (PD) based on handwriting analysis. It takes the extracted features from the handwriting samples and applies machine learning algorithms to categorize individuals as either PD patients or healthy controls. Once trained and validated, the model can effectively classify new handwriting samples, aiding in early diagnosis and monitoring of Parkinson's disease progression.

### 3.3.6 DECISION SUPPORT SYSTEM MODULE

The Decision Support System (DSS) plays a crucial role in assisting medical professionals by providing an automated, data-driven approach to Parkinson's disease (PD) diagnosis based on handwriting analysis. This module integrates the classification results from the machine learning model with a user-friendly interface that presents diagnostic insights in an interpretable format. It takes the predicted outcomes, along with confidence scores, and visualizes key handwriting features such as tremor intensity, speed variations, and stroke inconsistencies.

## 4. IMPLEMENTATION

### 4.1 HARDWARE IMPLEMENTATION

The Hardware Implementation of the Parkinson's disease diagnosis system involves integrating various devices to capture, process, and analyse handwriting samples efficiently. The primary hardware component is a digital tablet or smart pen, which records real-time handwriting data, including stroke pressure, writing speed, and tremor patterns. These devices are equipped with high-resolution touch sensors and pressure-sensitive technology to ensure precise data collection. In cases where digital tools are unavailable, scanners and cameras are used to digitize handwritten samples from paper for further processing.

A processing unit, such as a high-performance computer or embedded system is responsible for running machine learning algorithms and feature extraction processes. Cloud-based storage and computing resources may also be utilized to enable remote diagnosis and large-scale data processing.

Additionally, the system can integrate with wearable devices like smartwatches or motion sensors to capture additional movement-related biomarkers for enhanced accuracy. A user interface device, such as a touchscreen monitor or smartphone, allows doctors and patients to interact with the system, view diagnostic results, and track handwriting changes over time. By combining these hardware components, the system ensures efficient, accurate, and real-time Parkinson's disease diagnosis.



## 4.2 SOFTWARE IMPLEMENTATION

The Software Implementation of the Parkinson's disease diagnosis system using Python and Scikit-Learn (sklearn) involves multiple stages, including data preprocessing, feature extraction, model training, and classification. The system starts by collecting handwritten samples and digitizing them using OpenCV for image processing. Preprocessing techniques such as grayscale conversion, binarization, noise removal, and edge detection are applied to enhance the clarity of the handwriting data.

Using Scikit-Learn's feature extraction tools, static and dynamic handwriting features, including stroke width, writing speed, pressure variation, and tremor patterns, are extracted. These features are then fed into machine learning classifiers provided by Scikit-Learn, such as Support Vector Machine (SVM). The dataset is split into training and testing sets using `train_test_split`, and models are trained and evaluated using accuracy, precision, recall, and F1-score metrics.

```
input_data = (116.55600,592.03000,86.22800,0.00496,0.00004,0.00254,0.00263,0.00762,0.01660,0.15400,0.00820,
# changing input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)

#reshape the numpy array
input_data_resaped = input_data_as_numpy_array.reshape(1,-1)

#standardize the data
std_data = scaler.transform(input_data_resaped)

prediction = model.predict(std_data)
print(prediction)

if(prediction[0] == 0):
    print("The person does not have Parkinsons Disease")
else:
    print("The person has Parkinsons Disease")
```

FIG 3. SIMULATION OUTPUT

```
[ ] # accuracy score on training data
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(Y_train, X_train_prediction)

[ ] print('Accuracy Score of training data : ', training_data_accuracy)
➡ Accuracy Score of training data : 0.8846153846153846

[ ] # accuracy score on training data
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(Y_test, X_test_prediction)

▶ print('Accuracy Score of test data : ', test_data_accuracy)
➡ Accuracy Score of test data : 0.8717948717948718
```

FIG 4. ACCURACY SCORE

## 5. RESULT AND DISCUSSION

The system was tested on a dataset containing handwriting samples from both Parkinson's patients and healthy individuals, with a focus on Tamil and English scripts. After preprocessing and feature extraction, machine learning model such as Support Vector Machine (SVM). The results showed that SVM achieved the highest accuracy, around 92%, followed by Random Forest at 89%, indicating that dynamic handwriting features play a crucial role in distinguishing Parkinson's disease from normal handwriting patterns.

The discussion highlights the impact of key features such as tremor intensity, writing speed, and pressure variation, which were significantly different in Parkinson's patients compared to healthy individuals. Additionally, the study found that handwriting in Tamil script showed more noticeable irregularities due to its complex strokes and curves, making it a valuable indicator for PD diagnosis.

The system's real-time prediction capability and integration with a decision support system demonstrated its potential for early diagnosis and continuous monitoring. Future enhancements, including deep learning-based feature extraction and larger datasets, can further improve diagnostic accuracy and robustness. Overall, the results confirm that handwriting-based analysis, combined with machine learning, is a promising non-invasive tool for Parkinson's disease detection.

## 6. CONCLUSION

The diagnosis of Parkinson's disease (PD) using bilingual Tamil and English handwriting-based techniques presents a promising approach for early detection and monitoring of the disease. Handwriting analysis serves as a non-invasive, cost-effective, and efficient tool for identifying neuromotor impairments associated with PD. By examining variations in handwriting dynamics such as speed, pressure, tremors, and letter formation in both Tamil and English scripts, significant insights can be obtained into the severity of motor dysfunction.

The bilingual approach enhances the diagnostic accuracy as different scripts involve distinct motor and cognitive demands, providing a comprehensive evaluation of writing abnormalities. Advanced machine learning algorithms and deep learning models further aid in processing handwriting samples, enabling automated classification of PD symptoms with high precision. This research underscores the effectiveness of bilingual handwriting analysis as a supplementary diagnostic tool, potentially aiding clinicians in early detection, progression monitoring, and personalized treatment strategies for Parkinson's disease.

Future advancements integrating AI-driven models, larger datasets, and multimodal biomarker assessments could further enhance the reliability and applicability of this technique in real-world clinical settings.

## 7. REFERENCES

- [1] Jiao Meng, Xin Huo, Hui Zhao, Liming Zhang, Xun Wang, "Image-based Handwriting Analysis for Disease Diagnosis", 2022 41st Chinese Control Conference (CCC).
- [2] Theyazn H, H. Aldhyani, Abdullah H. Al-Nefaie, Deepika Koundal, "Modeling and diagnosis Parkinson disease by using hand drawing: deep learning model", 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET).
- [3] Mohamed Ziyadh M, Bavisetti Siva Avinash, K. Sreekumar, Sibi Amaran, "Integration of Voice and Handwritten Tests Analysis for Enhanced Detection of Parkinson's Disease using Machine Learning", 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC).
- [4] Mohamed Aghzal, Asmaa Mourhir, "Early Diagnosis of Parkinson's Disease based on Handwritten Patterns using Deep Learning", 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS).
- [5] Kazi Shaharair Sharif, Mohammed Majbah Uddin, Md Abubakkar, "NeuroSignal Precision: A Hierarchical Approach for Enhanced Insights in Parkinson's Disease Classification", 2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA).
- [6] Saeddin Kalash, Mohamad Ajaj, Nadine Abbas, Sirine Taleb, "Automated Early Detection of Parkinson's Disease Through Augmented Handwritten Patterns", 2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME).
- [7] Pranita Niraj Palsapure, B G Bhavana, M Jagadish, Kumbhar Trupti Ravikumar, "Detecting Early Signs of Parkinson's Disease: A Machine Learning-based Approach for Diagnostic Assistance", 2024 First International Conference on Software, Systems and Information Technology (SSITCON).
- [8] Amit Kumar Mishra, Vivek Deshpande, D. Anandhasilambarasan, Neelesh Singh, Vijayalakshmi G, Rekha M, "A Method for Identifying Parkinson's Disease from Spiral Images (Graphs) Created by Hand", 2024 Global Conference on Communications and Information Technologies (GCCIT).
- [9] Quoc Cuong Ngo, Nicole McConnell, Mohammad Abdul Motin, Barbara Polus, Arup Bhattacharya, Sanjay, "NeuroDiag: Software for Automated Diagnosis of Parkinson's Disease Using Handwriting", IEEE Journal of Translational Engineering in Health and Medicine ( Volume: 12): 2168-2372
- [10] Shanyu Dong, Jin Liu, Jianxin Wang, "Diagnosis of Parkinson's Disease Based on Hybrid Fusion Approach of Offline Handwriting Images", IEEE Signal Processing Letters ( Volume: 31)