Research on Engineering Structures & Materials
www.jresm.org
MIM Research Group

Research Article

Machine learning-based early detection of Parkinson's disease using handwriting and vocal features

Ashok R S [*,a], Anil Kumar D [b]

Dept. of ECE, BMS Institute of Technology and Management, Bengaluru, Affiliated to Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India

Article Info

Article History:

Received 22 Apr 2025

Accepted 24 Aug 2025

Keywords:

AlexNet;
DenseNet-121;
Deep learning in healthcare;
Handwriting analysis;
Jitter;
MFCC;
Parkinson's disease detection;
Speech-based classification;
Tremor;
VGG16

Abstract

This study investigates handwriting and speech patterns in individuals with Parkinson's disease (PD) using machine learning to enable early disease detection—a critical step for effective treatment. Handwriting analysis centers on motor components, such as spiral angle variation and wave amplitude, which reflect the impaired fine motor control characteristic of PD. Among deep learning models evaluated (ResNet, AlexNet, DenseNet, and VGG16), the DenseNet-121 model achieved the highest accuracy of 85.17% for classifying motor control differences. Voice analysis targets non-motor symptoms, focusing on speech disturbances linked to tremors and muscle rigidity. Machine learning classifiers (SVM, KNN, MLP, XGBoost, Logistic Regressor, and Random Tree) were implemented, with SVM demonstrating the best performance by reaching an accuracy of 89.74% alongside strong precision and recall. Combining handwriting and speech analysis offers a more comprehensive and effective PD diagnosis than conventional clinical approaches, facilitating prompt intervention for improved patient care.

1. Introduction

Parkinson's disease is second most common neurodegenerative disorders that generally showcases symptoms associated with movement [1]. The loss of dopamine-producing neurons in the substantia nigra results in this condition. The substantia nigra is essentially a brain region that controls movement and coordination. An important neurotransmitter, its lack leads to tremors, muscle rigidity, bradykinesia, and postural instability among the characteristic motor symptoms of PD. Even though PD typically occurs in those aged 60 years or more, there are instances of early onset. In those affected, progression also varies; however, it is chronic-progressive, meaning it worsens gradually.

It is not clear what causes Parkinson's disease, although studies have shown that genetic and environmental factors contribute equally to it. Exposure to pesticides, herbicides, and industrial chemicals has also been linked to an increased incidence of Parkinson's disease. Furthermore, head trauma or injury have been associated in some reports with increased susceptibility to PD. It is a significant risk factor since the incidence of this disease does increase with age [2].

Motor symptoms include tremors, muscle rigidity, bradykinesia, balance, and coordination problems. The symptoms generally occur asymmetrically - they affect one part of the body more than the other. Patients develop difficulties in walking, shuffling gait, and stooping posture as

*Corresponding author: ashokchnagar@gmail.com
[a]orcid.org/0009-0004-0033-0410; [b]orcid.org/0000-0003-4711-8359
DOI: http://dx.doi.org/10.17515/resm2025-835ml0422rs

symptoms progress in the course of the disease. In addition to motor symptoms, many non-motor symptoms indicating Parkinson's disease, include insomnia, depression, anxiety, cognitive decline, and autonomic dysfunction, such as constipation or low blood pressure. In some cases, the non-motor symptoms begin before the beginning of the typical symptoms, which can lead to challenging diagnosis. While there is no cure for Parkinson’s disease, a combination of medications like levodopa, therapies such as DBS, and supportive treatments including physiotherapy, speech therapy, and healthy lifestyle choices can help manage symptoms and improve quality of life [3].

One of the promising avenues of early diagnosis in motor impairment may be best understood by analyzing handwriting, particularly with regards to spiral and wave drawings. Since fine motor skills are lost over a long period of time in a person suffering from PD due to the progression of bradykinesia, rigidity, and tremors, such loss can sometimes be translated through a person's handwriting. Both these clinical assessments have merit. Spirals produced by such patients are mostly irregular, shaky, or may have discontinuities in them because of the incompetence of holding fluid and continuous movements. New advances in machine learning enable the quantification and analysis of these drawings so that at a stage when signs may still be minimal through other clinical ways of expressing them, motor abnormalities become detectable.

A notable aspect of handwriting analysis, by spiral and wave drawings, is that it can be a source of information about the motor function of the patient with Parkinson's disease. Tremors, rigidity, and bradykinesia, all these features characterize most tasks, including writing and make them hard to do for PD patients. Spiro-graphic and wavering movements are also typically used during clinical tests in assessing these motor impairments. Deviations in the smooth, continuous lines that characterize spiral drawings, such as unevenness, tremors, or irregularities in the shape of the spirals, may indicate motor dysfunction. This kind of drawing may provide an essentially non-invasive and simple way to detect early signs of PD, since motor deterioration often first manifests itself in subtle changes in handwriting long before any more overt symptoms, such as walking difficulty, occur. These drawings can be automatically analyzed through machine learning algorithms to set quantitative values for the severity of motor impairment. Thus, an objective monitoring system would be generated for the progression of the disease [4].

In parallel, voice analysis is a highly relevant tool for diagnosing the non-motor functions due to PD. The alterations in voice include decreased volume, hoarseness, monotonization, and problems with articulation. Many of these manifestations occur because of the involvement of muscles responsible for speech in PD patients. These vocal changes usually occur at an early stage of the disease and can be captured based on acoustic features such as jitter, shimmer, and variations in pitch. Machine learning models can identify obscure patterns that suggest early signs of PD based on these parameters. It is thus possible to evaluate PD at an earlier stage and during different phases by combining motor assessments through handwriting analysis with non-motor evaluations via voice analysis. This multi-faceted approach stands a better chance of being diagnosed effectively in time and subsequently leads to improved patient outcomes.

Clinical examination, neuroimaging, and biomarker assessment are other methods for diagnosing PD apart from handwriting and voice analysis. Clinical examinations mainly employ some standardized scoring systems which rate the severity of the illness, a good example being the Unified PD Rating Scale (UPDRS) which assesses the motor and non-motor symptoms after interviewing the patient and performing physical examination. Neuroimaging consists of MRI and PET scanning and it serves to illustrate the structure and activity of the brain, particularly the changes in the dopamine transporter which is observed at the early stages of the disease [5].

Machine learning and artificial intelligence have made it possible to analyze hand-writing and voice data with possibilities that were previously unimaginable, setting up copious amounts of processing to identify early signs and patterns related to PD. These advances help with remote monitoring by sensors and mobile health applications that provide steady, continuous feedback and timely alerts, allowing for further timely interventions, which make a great deal of difference in quality of life. More people will be able to get the help they need early with diagnoses.The novelty proposed study is thus on new methods of handwriting and voice analysis that rely on machine learning and data analysis techniques toward comprehensive approach early diagnosis for PD.

Combining the strengths of both modalities, this can progress toward a multi-faceted analytical framework toward enhanced specificity and sensitivity in diagnostics, paving the way for effective monitoring and subsequent better management of the disease.

2. Related Work

A review that focuses on the idea of revolutionizing early detection of various health conditions based on the machine learning techniques and its applications. The traditional methods of disease detection are quite challenging with high costs and issues of accessibility. Thus, new and innovative methods come with the urgency of need. Using non-invasive and cost-effective approaches, including voice analysis and handwriting analysis combined with machine learning algorithms might help us identify pertinent patterns of interest and enhance classification accuracy. This review demonstrates some of the recent advances made in feature selection and classification methods that have achieved efficacies in identifying health conditions. In this regard, this review charts out to date as well as future directions in the implementation of machine learning-based approaches in the diagnosis process to continue to empower enhanced healthcare outcomes and efficiency.

R. K. Sharma et. al. [6] proposed that PD, being one of the neurological disorders characterized by the deficiency of dopamine neurons, plays an essential role in managing body movements. They are people suffering from PD who usually struggle to perform day-to-day activities and have disturbed vocal fold movements. The current study aims to discuss the potential use of voice analysis for remote and accurate early-stage disease diagnosis without increased costs. The voice features dataset of 23 showed a good significance of 15 features, namely jitter, shimmer, harmonic-to-noise ratio, DFA, Spread1, and PPE that all refer to tremor variations of the vocal box muscles associated with PD. Different classifiers have been adopted in this work to look for the most suitable one for the purpose of detection. Interestingly, it produced support vector classifiers that surpassed other discriminators by yielding an impressive accuracy of 96%. In addition, several kinds of different neural network classifiers were compared by their transfer functions in order to judge classifiers for this task.

I. Nissar et. al. [7] proposed an article that examines the application of machine learning for the identification of PD through voice analysis. In this regard, the study investigates the effect of type of feature selection, which would either be Mel-Frequency Cepstral Coefficients (MFCC) or Transformed Wavelet Transform (TQWT), on the efficiency of the system. It compares various machine learning models, such as Logistic Regression, Naive Bayes, KNN, Random Forest, Decision Tree, SVM, MLP, XGBoost, and the suitability of their application in PD detection. Techniques like minimum-Redundancy and Maximum-Relevance (mRMR) and Recursive Feature Elimination (RFE) are applied to feature selection. The XGBoost model with mRMR feature selection technique in combination achieved the highest performance, with the accuracy result being at 95.39%, with both MFCC and TQWT features used, and precision, recall, and F1-score at 0.95. These results, therefore, strongly support the use of the XGBoost model for PD voice sample-based detection as the effectiveness of mRMR feature selection technique used improves the model performance.

S. V. T. Dao et. al., [8] approach to finding PD from voice recordings, as this disorder is critical, which affects about ten million people worldwide, significantly affecting their daily life. Conventional detection methods usually utilize expensive and inaccessible techniques, so voice analysis becomes a promising non-invasive and cost-effective alternative for earlier detection. To summarize, the work identifies significant vocal patterns by using GWO for feature selection and improves classification performance using the LGBM algorithm. The proposed model has demonstrated competitive results, with accuracies of 0.878 K-NN, 0.866 SVM, 0.795 Decision Trees, and 0.894 LGBM for classifications between individuals diagnosed with PD or healthy. It has the potential of timely treatment recommendation and may lead to better patient outcomes so further development and implementation is expected in real health care in practice.

A. Suppa et al., [9] have done an Investigation on changes in voice in patients suffering from PD, hypokinetic dysarthria, using machine learning algorithms to analyze recordings from a cohort of 115 PD patients with a mean age of 68.2 years. A total of 57 early-stage PD patients were untreated with L-Dopa and 58 mid-advanced-stage PD patients were chronically treated with L-Dopa. For comparison, this study also recruited a healthy control group comprising 108 age-matched healthy subjects with a mean age of 60.2 years. Ubiquitous voice assessments have been carried out with Unified PD Rating Scale and Voice Handicap Index with a support vector machine classifier applied to audio recordings. Results demonstrate that voice abnormalities exist early in the disease process but do systematically deteriorate through disease progression; L-Dopa can provide improvement but does not return voice quality towards normal. High accuracy was achieved between healthy and PD patients at all stages, as well as among patients OFF and ON L-Dopa therapy. A new score obtained from machine learning also established significant clinical-instrumental correlations that may define this biomarker for PD. Overall, the results show the effective role of machine learning for monitoring the severity of the disease and treatment effects on voice parameters in patients with PD.

K. P. Swain et. al., [10] aims to derive and establish the feasibility of applying machine learning algorithms in the early diagnosis of PD by using an analysis of voice recordings toward the development of a non-invasive, reliable, and viable diagnostic method that would better address early interventions and management. The application used a dataset of 195 voice samples with 23 features after data preprocessing and balancing. The KNN model showed higher precision of (0.96-1.00), recall of (0.97-1.00), and F-scores of (0.98-0.99) while getting an overall accuracy of 0.98 on 59 samples. Such research serves to underscore the potential that the application of machine learning might reveal for diagnosing PD and advocates the KNN model as a promising tool for early diagnosis, thus promoting the application of machine learning techniques in more robust ways in health care.

O. P. Neto, et. al. [11] presents a paper evaluates the performance of voice analysis together with machine learning techniques in determining cases of PD. The analysis is performed on voice data, specifically the phonation of the vowel 'a', from three datasets comprising 432 participants-278 PD patients. Four machine learning models such as Artificial Neural Networks (ANN), Random Forest (RF), Gradient Boosting (GB), and Support Vector Machine (SVM) with two ensemble methods namely soft voting classifier (EVC), and stacking method (ESM) were applied. The performance of the models was tested over 50 iterations with different data splits and 10-fold cross-validation, and the goodness of fit between the models was compared using one-way ANOVA with Bonferroni post hoc corrections. The results demonstrated that among the three, ESM, SVM, and GB were ranked to be the best performers since their scores for all the metrics showed to be very high, though it was beaten by heterogeneous data and was dogged by variable selection challenges: accuracy, sensitivity, specificity, precision, F1 score, and ROC AUC. The promise indicated in the integration of ML techniques with voice analysis for early diagnosis of PD points out the importance of using multi-source data and large sample size for developing the validity, reliability, and generalizability of the results. In this regard, speech-language pathologists would be motivated by the findings given that they come with tools designed to improve, refine, and fine-tune diagnostic processes and facilitate early intervention within clinical settings.

S Aich et. al., [12] brought a paper, that will be described, why there is a need for specific biomarkers for clinical decision systems, especially for patients with PD. It will be noticed that handwriting impairment correlates with disease severity: generally, PD patients write at reduced speed and pressure. The proposed system will analyze the spiral and wave drawing patterns in patients diagnosed with PD and healthy individuals. For the evaluation of drawing patterns, the system has recourse to two different CNNs. Predictive values obtained from both networks were combined by an ensemble voting method along with a metal classifier to increase the accuracy of the prediction. Training was performed based on data taken from 55 patients and achieved a remarkable overall accuracy of 93.3%, and also a mean recall of 94%, a mean precision of 93.5%, and a mean F1 score of 93.94%.

S C S kar et. al., [13] proposed a research article regarding PD, a prevalent neurodegenerative disease that often presents with the characteristics of an impaired motor system, including a lack of coordination, a resting tremor, and rigidity. It is significantly characterized by issues concerning handwriting, specifically a condition called Micrographia. Machine learning-based analysis of Static

Spiral Tests (SST) of PD patients with the K-Nearest Neighbors (KNN) algorithm for the two-class categorization of the spiral drawings from two classes: healthy individuals and PD patients. It is expensive as well as time-consuming to clinically diagnose the PD; hence, the automation of SST analysis may bring efficiency in the diagnosis and monitoring of neurological conditions. It classifies the patients versus healthy subjects with an accuracy of 96.07% and has reached accuracies greater than 90% on a separate validation set. A web application called "PD Detector" is also presented for early detection of PD based on the proposed model.

M Singh et. al. [14], presented a paper for detection of PD using spiral sketching from analysis by Convolutional Neural Networks. The basic idea here would be to classify a spiral drawing as healthy or indicative of PD. Healthy subjects usually generate spirals that look very much like regular shapes, while those with PD have distorted spirals due to slow movement and poor coordination between their hands and the brain. The experiments prove that in this analysis, the CNNs reach a classification accuracy of 83.6%; spiral sketching, therefore, can prove to be a highly effective diagnostic tool for PD.

Z A Shaikh et. al., [15] in the research uses biometrics analysis for the detection of PD. It provides the effectiveness of handwriting impairment in correlating with the severity of disease. The proposed system uses a spiral and wave line drawing pattern analysis through the application of machine learning algorithms. In this study, two CNN models have developed and achieved higher values of accuracy in the Spirals Model at 98%, Wave Model at 84.61%, etc. Further, the treatment adopted an admin panel in Django, which was helpful for efficient management and organization of results obtained from the diagnosis. In general, the research has provided a viable avenue for early diagnosis and intervention of PD since it enables a smooth management of results with some appropriate model predictions.

Y. Huang et. al., [16] showed Parkinson's is a progressive brain disorder that affects millions due to the degeneration of the dopamine producing brain cells, which affects movement, balance, and posture. More importantly, it speaks about an early diagnosis for betterment in the quality of life of the patients. In this paper, therefore, I outline a handwriting-based prediction approach by combining a cosine annealing scheduler with deep transfer learning. Using the NIATS dataset, which is a collection of both PD patients' handwriting and healthy individuals' handwriting samples, the paper compares the performance of six different models: VGG16, VGG19, ResNet18, ResNet50, ResNet101, and ViT. These models vary in terms of accuracy, precision, and F1 scores: the proposed method combined with the VGG19 model yields the highest average accuracy of 96.67%.

F. Mercaldo et. al., [17] jagged difficulties associated with the diagnosis of PD; there is no specific test and up to 90% of untrained individuals end up misdiagnosing. This paper specifically looks at the spiral drawing test-a clinical test applied to assess fine motor ability and hand-eye coordination and the presence of tremors among patients with neurological disorders. This experiment traces a participant in a spiral pattern. Any anomaly in a participant's tracing movement may indicate to healthcare professionals whether the individual suffers from what is commonly referred to as PD or any of the other forms of essential tremors. This proposed study allows for the spiral drawing test, which can be analyzed through an automatic method by using two Convolutional Neural Networks: DenseNet and ResNet50. Results reveals that the technique was highly accurate up to the mark, which was 96%, during the test performed on 3,991 spiral drawing tests, thereby proving to be an efficient method. The approach also features a visualization tool that displays relevant areas in the test image to the model regarding its prediction of PD, thereby giving some idea of the decision-making process of the model.

L. Lonini et al., [18] illustrates the feasibility of applying machine learning algorithms on data measured by soft wearable sensors to accurately and automatically detect the occurrence of symptoms like PD and monitor the progression of the disease in patients. The researchers note that annotated data from clinical experts is expensive and laborious to obtain; therefore, the researchers collected movement data with six flexible wearable sensors worn by 20 individuals with PD across multiple clinical assessments on the same day and again two weeks later. Participants performed 13 common tasks, including walking and typing, while clinicians rated severity of symptoms such as bradykinesia and tremor. The researchers then trained convolutional neural networks and statistical ensembles to detect these features from the collected data. Interestingly, one sensor on the back of the hand was found to be sufficient to detect bradykinesia in the upper extremity and, once more for the tremor case, no extra benefit beyond using sensors on both sides. In contrast, training the model on more people, though enhancing the performance of the model, had no strong effect on detection accuracy when assessing the same individuals over days. Overall, the results imply that people suffering from PD symptoms can be well differentiated at different times with the help of datasets that typically encompass diverse cases of individuals.

Zubiena et al., [19] proposed a study to used dynamic posturography with wearable sensors for the early detection of balance dysfunction in sub-clinical patients with PD. Although the method used herein is highly sensitive, it has several limitations because the analysis process is quite complex and therefore not feasible in routine clinical practice. The study group used machine learning algorithms in distinguishing patients suffering from PD and healthy control subjects as well as distinguishing between the OFF and ON states of dopaminergic therapy. Data were obtained from 20 PD patients and 15 healthy subjects. It tested 52 classifiers based on decision tree, K-nearest neighbor algorithms, support vector machine algorithms, and artificial neural network algorithms. Twenty-one classifiers met the inclusion criteria, and Fine K-Nearest Neighbor proved to be the most efficient classifier for PD patients irrespective of their ON or OFF state condition. However, none of the classifiers could identify the ON vs OFF states. Altogether, the results suggest that, using machine learning, automated kinematic data analysis could be useful for the early diagnosis of balance disorders in PD patients.

M. G. Krokidis et al., [20] wrote a research article to discusses PD as a chronic progressive neurodegenerative disorder wherein dysfunction of dopaminergic neurons and dopamine deficiency along with the formation of Lewy body protein particles result. It has emerged that sensor-based platforms have now become very valuable tools in clinical practice, through which many biological signals can be screened simultaneously as well as a large number of biomarkers may be promptly taken for diagnosis and prognosis. Integration of machine learning into medical systems provides an opportunity to optimize the collection of data and to improve the prediction of the disease through the classification of its symptoms, thus supporting data-driven clinical decisions. The paper deals with the current state of sensor-based approaches in PD diagnostics and subsequently discusses the used of ensemble techniques in conjunction with sensor data to develop machine learning models for personal risk prediction.

Previous studies on PD detection have focused either on handwriting or voice analysis as independent indicators of the disease. Handwriting-based detection often utilizes spiral drawing tasks, leveraging features like tremor frequency, stroke smoothness, and angle variation. Earlier works have applied traditional image processing techniques and shallow classifiers such as SVMs or Decision Trees, with modest classification accuracy around 70–80%. Deep learning approaches like CNNs have more recently been employed, showing improved performance due to their ability to automatically learn intricate motor pattern features. However, most of these studies relied on a single type of deep network and did not explore a comparative analysis among multiple deep architectures. Additionally, integration of voice analysis into PD diagnosis is less common in earlier work, though some efforts using MFCC (Mel-frequency cepstral coefficients) and basic acoustic features with classifiers like k-NN and logistic regression have shown promise, often achieving accuracies in the 80–85% range.

In contrast to these prior studies, the present work innovatively integrates both handwriting and speech analysis, thus capturing both motor and non-motor symptoms of PD for a more holistic diagnostic approach. It also distinguishes itself by conducting comparative evaluations across a broad range of machine learning and deep learning models. DenseNet-121 achieved the best handwriting-based classification with an accuracy of 85.17%, while SVM led the voice analysis with 89.74% accuracy, surpassing results in most previous studies. This dual-modality approach and model comparison offer a more comprehensive and accurate PD detection framework, potentially enhancing early diagnosis efficacy beyond what earlier single-modality methods could achieve.

3. Problem Statement

The growing need for accurate, non-invasive, and scalable methods for early detection of PD highlights significant gaps in conventional diagnostic approaches, which often depend on subjective clinical assessments. These traditional methods can lead to misdiagnosis or delays in identifying the disease, ultimately affecting patient outcomes. This work aims to integrate real-time data from voice analysis, handwriting recognition through wave and spiral patterns, and advanced machine learning algorithms. By leveraging these innovative techniques, the objective is to develop a robust system that enables timely diagnosis and intervention, enhancing daily well-being for those at risk of PD. This approach not only addresses the limitations of existing diagnostic practices but also paves the way for more precise, objective assessments that can facilitate earlier and more effective treatment strategies.

4. Objective

The primary aim of this research is to propose a holistic diagnostic framework toward the early detection of PD. This could be established by providing an integration of voice analysis with handwriting evaluation. The proposed system will be developed with the combination of advanced algorithms and specific feature extraction from recordings of voice samples and hand-drawn spiral patterns to provide timely insights for the healthcare professional. Such a system can allow better applications of early intervention strategies while ensuring better patient outcomes without sacrificing scalability and robustness across diverse patient demographics.

One more focus of the research was to study advanced feature extraction techniques by utilizing deep learning models that include CNNs and RNNs. These models are actually applied to capture the very complex patterns in voice and speech data towards improving the diagnostic precision. The signal processing algorithms improve the incorporation of machine learning approaches, aiming to detect the features strongly associated with the progression and severity of PD. The directions are of course toward improving methods for early detection.

The central themes of this research involve the development and optimization of algorithms extracting key features from voice recordings of patients with PD. Parameters or functions corresponding to fundamental frequency variations, intensity fluctuations, and prosodic elements are included in the extracted set of features. These will refine the selection of features, particularly in further improving the sensitivity and specificity of diagnostic models at very low computational costs without losing any aspect of performance. The models are validated using proprietary and publicly accessible voice datasets of proven reliability in performance.

This study explores the used of machine learning models to analyze and interpret handwriting patterns, especially in cases such as spiral and wave drawings captured through digital means. This research will identify micrographia and tremor-induced irregularities typical among Parkinson's patients. The study used annotated handwriting datasets for training to better the precision in determining which patterns are correlated with the actual degree of disease severity with a view to developing diagnostic tools in handwriting analysis.

The developed algorithms are validated with clinical datasets of voice recordings, handwriting samples, and medical assessments from Parkinson's patients. Severe cross-validation against long-term data and various patient groups is performed to test whether the algorithms classify correctly between patients who have PD and healthy controls. This validation is important for establishing reliability and clinical applicability of the diagnostic models.

Translation of findings into clinical practice is similarly provided by the focus of the research. In collaboration with medical institutions, diagnostic testing will be conducted with the utilised of a pilot test for the platform in real-world settings that streamline the diagnosis and encourage early intervention, aligning research with applied clinical work for patients' betterment through technology solutions in neurology.A suitable test and validation protocol, including cross-validation and external testing with independent datasets will be used to test the algorithms comprehensively and in-depth in order to evaluate them. This will ensure the reliability and effectiveness of the developed diagnostic tools under most conditions. In this regard, several metrics like accuracy, sensitivity, specificity, and robustness to noise will be tested for confirmation of the fact that algorithms perform well in real clinical environments.

5. Methodology

This study adopts a comprehensive and integrated approach toward the design and validation of an intelligent system for the early-stage detection of PD . By leveraging multimodal data sources, including speech signals and handwriting patterns, the proposed framework aims to enable robust early diagnosis. The methodology involves systematic feature extraction from both modalities, capturing clinically relevant biomarkers such as vocal perturbations (e.g., jitter, shimmer, HNR) and handwriting dynamics (e.g., pen pressure, drawing velocity, and inter-stroke features like CISP). These features are then processed and input into a set of supervised machine learning classifiers, which are trained to discriminate between healthy control subjects and PD patients.

To ensure robustness and generalizability, the models were evaluated using k-fold cross-validation as well as external validation on clinically annotated datasets. Performance metrics such as accuracy, sensitivity (recall), specificity, precision, and F1-score were employed to quantify model efficacy. Additionally, a pilot deployment was conducted in collaboration with clinical institutions to assess the system's real-world applicability and operational feasibility in a healthcare setting. The machine learning algorithms explored include SVM, Random Forests (RF), k-Nearest Neighbors (KNN), Logistic Regression (LR), Multi-Layer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost), with hyperparameter tuning performed to optimize predictive performance.

5.1. Spiral and Wave Analysis for Handwriting Assessment

5.1.1. ResNet

ResNet (Residual Network) - a novel architecture that allows the handling of the challenges of training very deep networks by providing a new architecture where skip connections allow to learn residual functions, rather than mappings. It can be written mathematically as mentioned in equation (1):

$$H(x) = F(x, \{Wi\}) + x \tag{1}$$

The output of the residual block is $H(x)$. The function learnt by the network is called the residual function, given as $F(x, \{Wi\})$. The input to the network is denoted by x. The weights of the layers in the block are represented as Wi. The skip connections introduced in ResNet have led to effective flow back of gradients when backpropagation occurs and have improved against the vanishing gradient problem, allowing training network as deep as hundreds or even thousands of layers without degradation of performance. The result of this was an effective advancement in computer vision fields from notable improvements in tasks related to image classification and object detection.

5.1.2. DenseNet

DenseNet, or in other words, Densely Connected Convolutional Network, is a deep learning architecture that improves the feature propagation while reducing the number of parameters in deep networks. Its feature is founded on the idea of dense connections in which every layer receives input from all previously connected layers; hence, information flows freely throughout the network. This may mathematically be represented as mentioned in equation (2)

$$H_l \ = \ H_{l-1} + F(H_{l-1}, W_l) \tag{2}$$

In this equation, H_l is the output of layer l, H_{l-1} is the output from the previous layer, and $F(H_{l-1}, W_l)$ represents the transformation applied at layer l with weights W_l .Dense connections enable the network to reuse features, support the flow of gradients, and remove redundancy, thus improving efficiency and performance. DenseNet is a structure that greatly reduces the parameters in comparison with the traditional architectures and maintains high accuracy on tasks of image classification and object detection.

5.1.3. AlexNet

AlexNet is a pioneering deep learning architecture which considerably advances the field of computer vision, especially the computer vision field in image classification tasks, by connecting convolutional layers followed by max-pooling layers to capture spatial hierarchies in images and reduce dimensions. The architecture can be mathematically represented by the convolution operation as shown in equation (3).

$$Y = f\,(W3\ \mathrm{ReLU}(\ W2\ \ \mathrm{maxpool}\ (\ W1\ x))) \tag{3}$$

In this equation, where W1, W2, W3 are the weight matrices, and ReLU is the activation function used to introduce non -linearity. AlexNet incorporates five convolutional layers followed by three fully connected layers and employs dropout for regularization along with data augmentation to improve generalization. Such architecture, trained on the ImageNet dataset, resulted in breakthrough performance in the 2012 ILSVRC.

5.1.4. VGG16 Model

The VGG16 is a very simple architecture of deep convolutional neural networks. It demonstrates effectiveness in a straightforward manner within the image classification tasks. The utilisation of small convolutional filters stacked in increasing depth helps increase complexity for learned patterns while remaining computationally efficient. In mathematics, the operation of convolution can be written as mentioned in equation (4):

$$Y_l = \ \ maxpool\ (ReLU(W_l\ \cdot X_l)) \tag{4}$$

The VGG16 model is a learning model with a 16-layer depth and learnable parameters, including 13 convolutional layers and three tail fully connected layers. Max-pooling layers are applied in addition to reduce the spatial dimension, giving a deeper representation without extra computations. VGG16 is built with a structured approach and depth, and it will go on to become the fundamental model for many architectures in computer vision, especially through feature extraction and transfer learning.

5.2. Speech Analysis for Voice Feature Assessment

5.2.1. Support Vector Machine

A Support Vector Machine is a supervised learning model with application both in the classification and in the regression tasks. SVM theory constructs a hyperplane that separates data points from different classes in high-dimensional space. The typical mathematical expression behind SVM is finding the maximum-margin hyperplane between support vectors, which are the closest data points from different classes. It could be formulated as an optimization problem which is shown in (5).

$$min\ \tfrac{1}{2}\ ||W||2\ \ \text{subject to}\ \ Yi\ (\ W \cdot\ Xi - b) \geq 1\ \forall\ i \tag{5}$$

Here, W is the weight vector, b is the bias, and Yi are the class labels. The most important feature of SVM is the made use of a trick called 'kernel trick' to transform an input into some higher dimension and then used this for generating complex decision boundaries. It is the capability thus offered that enables SVM to handle any nonlinear data with ease, and so it has found extensive utilisation as a smart tool for classification.

5.2.2. K - Nearest Neighbor

K - Nearest Neighbors is a very simple and yet one of the very effective algorithms used for supervised learning classification and regression tasks. The basic idea of the KNN algorithm is to classify a data point based upon the majority class among its k closest neighbors in the feature space. Mathematically, to get the distance of all other points to a query point x in the training dataset, as mentioned in the equation (6):

$$D(x,xi) = \sqrt{\sum_{j=1}^{n} (x_i - x_{ij})^2} \quad (6)$$

where D is the distance between two points in data, where x is the training sample, x_j is the j-th feature of the query point, and x_{ij} is the j-th feature of the i-th training sample. The algorithm selects the KNN and assigns the class based on the majority vote among them. The most prominent characteristic of KNN, however is that it is simple and effective, even for the multi-class classification problems as it requires no assumption about its underlying data distribution. KNN is also adaptive, and therefore capable of dealing with dynamic data where the class labels are not static but may change over time.

5.2.3. Logistic Regressor

Logistic regression is among the statistical methods applied in binary classification, whose intention is to model the probability of an input belonging to some specific class. Unlike linear regression, logistic regression creates the probability of a binary occurrence rather than a continuous value through the used of the logistic function or sigmoid function. The mathematical expression for logistic regression is given in equation 7.

$$P(y = 1|x) = \frac{1}{1 + exp(-WTx)} \quad (7)$$

Here, W is the weight vector. The characteristic of logistic regression that is considered the most important is interpretability. The coefficients are understandable as the contribution of each feature in terms of the log-odds of the outcome, and that's why it is widely used in medicine, finance, social sciences, for instance, for modeling binary outcomes. It is also possible to consume this for multi-class classification with one-vs-all or softmax regression techniques.

5.2.4. Random Forest Classifier

A Random Forest Classifier is an ensemble learning algorithm, wherein several decision trees are built together by the help of bagging and feature selection to make it better for classification and avoid overfitting. It does function on the principle of many decision trees being built in training, making the output class the result of individual trees' majority vote. Each tree will be trained on a random subset of the features and data to introduce diversity between the trees and improve generalization. The classification decision for any input x is an average of all the predictions from T decision trees as in equation (8).

$$\hat{y} = mode(\hat{y}1(x), \hat{y}2(x), \ldots, \hat{y}r(x)) \quad (8)$$

where ŷt(x) is the prediction of the t-th decision tree for input x, and the final prediction ŷ is the most common prediction (mode) across all trees. The Random Forest Classifier most importantly handles much higher dimensionality even when the data size is big, with robust predictions having lower overfitting chances as against the independent decision tree. It's also versatile in that it can handle the classification and regression tasks, even handle missing data and maintain accuracy up to its full potential even when most of the data is missing.

5.2.5. Neural Network MLP

A multi-layer perceptron is a feed-forward artificial neural network used to classify objects and to solve regression problems. The MLP consists of an input layer, one or more hidden layers, and an output layer. In each layer, there are connections between neurons and every neuron in the next layer with weights on the connections; data travels forward through the network. Each neuron computes its output by passing a weighted sum of inputs through a non-linear activation function, for example, the ReLU or sigmoid functions. Mathematically, a neuron's output in the hidden layer will take the form as in equation (9).

$$hj = f \sum_{i=0}^{n} Wji\, Xi \; + \; bj \tag{9}$$

where hj is the output of the j-th hidden neuron,Xi represents inputs, Wji is the weight from input Xi to the jth neuron, bj is bias, and f is the activation function. The model produces predictions based upon the output in the hidden layer of its output layer and then back-propagated with the aid of a learning algorithm in order to minimize the errors between predicted and actual outputs. The most appealing feature of MLP is that it allows learning of complex, nonlinear input/output relationships due to a number of layers of neurons and activation function used. MLPs are, therefore, very flexible and capable of modeling intricate patterns.

5.2.6. XGBoost

XGBoost is a mature, highly-efficient machine learning algorithm in the class of gradient-boosting methods designed for both classification and regression problems. It constructs an additive model in a forward fashion where each subsequent tree attempt to reduce the error of previous trees. At each step the model minimizes the objective function -- that's the aggregation of the loss function with the assessment of its predictive error plus a regularization term, controlling the level of complexity of the model as in equation (10)

$$L(\emptyset) = \sum_{i=1}^{n} l(yi, \hat{y}i) \;\; + \sum_{k=1}^{K} \alpha(Fk) \tag{10}$$

Of course, the most significant feature of the algorithm is its ability to scale and be effective in performance with optimizations, including constructing trees parallelly, regularization, handling missing values, and support for sparse data. To explain briefly, those features make XGBoost really effective for those real applications on big, complex datasets.

6. Implementation

Through this work, we focus on PD, which is an incurable neurodegenerative disorder. It is primarily related with movement regulation and includes symptoms such as tremors, stiffness of muscles, and imbalance. Traditionally, diagnostics have been mainly clinical and other scanning techniques that were subjective in nature and often did not reach the early stages of the disease. However, current technology would provide novel solutions to better enhance the improvement of accuracy and efficiency in diagnosing and monitoring PD. One of these innovations involves the utilistion of deep learning models to analyze both motor and non-motor parameters, thereby gaining a better understanding of the impact of the disease.

The number of deep architectures involved in the analyses of motor parameters for the spiral and wave drawing tasks-cases proposed to be used by the proposed system is quite a number, which includes ResNet-34, ResNet-50, DenseNet-121, DenseNet-169, VGG-16 and AlexNet. The models learned hand-drawn spirals and waves from these datasets and capture the peculiar character of motor control both in the affected and unaffected individuals. These drawings will, therefore, be useful in differentiating PD patients from healthy individuals so as to provide a reliable and objective way of monitoring the evolution of motor symptoms. Their motor impairments would, therefore, be well assessed accurately and continuously with this integration of state-of-the-art techniques in deep learning.

Apart from the above motor analysis, the framework has introduced voice signal analysis in considering the non-motor symptoms of PD. This system extracts features from speech and tremor patterns in voice recordings using machine learning classifiers like logistic regression, K-nearest neighbors, support vector machines, the Random forests classifier, XGBoost, and Neural Network MLP, in order to map features to separate affected and unaffected patients. Hence, this would provide a more comprehensive approach toward PD diagnosis and monitoring, integrating insights from both motor and non-motor parameters to promote early detection and better strategies toward patient management.

6.2. Dataset Acquisition

6.2.1. Drawing Dataset - The Michael J. Fox Foundation

The Michael J. Fox Foundation provided the dataset for the analysis of spiral and wave drawings. In PD, more severe disease has drawn slower and lower pen pressure with higher severity of PD and 0.4 at Severity Level (SL). Therefore, further work is needed on such features that are more accurately correlated with SL. This paper presents Correlation of Inter-Stroke Pressure (CISP) as a novel feature of PD severity.

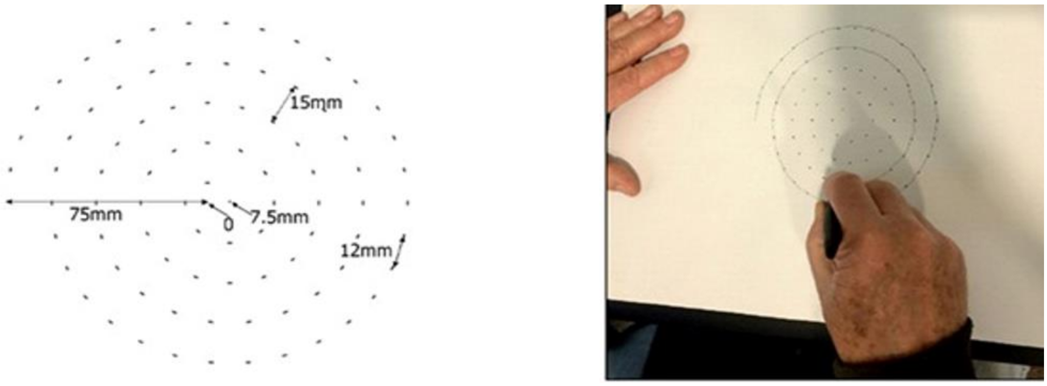
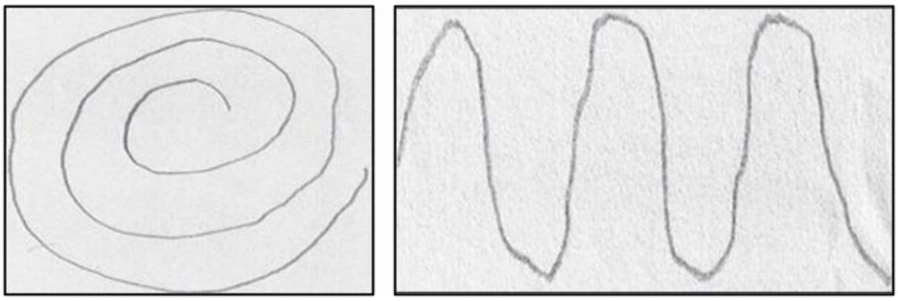
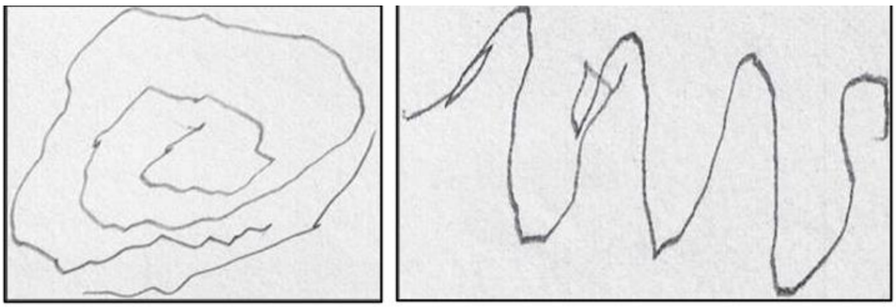


Fig. 1. Spiral and wave drawing dataset



(a)



(b)

Fig. 2. Spiral and wave drawing dataset classification (a) healthy data set and(b) affected data set

The total of twenty-seven PD patients and 28 controls participated in the study. All the participants performed guided Archimedean spirals on an A3 sheet to get their UPDRS score. The calculated features in this study are speed, pen pressure, and CISP (Complexity Index of Spiral Precision) with every feature for disease severity assessment. Results have presented correlation coefficients of -0.415, -0.584, and -0.641 between the speed, pen pressure, and CISP respectively, and PD severity. Mann-Whitney U test showed a significant difference in the PD patients versus controls. Non-parametric Kruskal-Wallis test proved a statistically significant difference among the levels of PD severity classified with high sensitivity by CISP – SL-1 and SL-3. Hence, the results exhibit that CISP can differ between healthy controls and PD patients, while it can classify between SL-1 and SL-3 of

PD and is not too apt at classifying between PD SL-2. The figure 1 shows the representation of handwriting dataset through spiral image drawings. Figure 2 showcases the classifications.

6.2.2. Voice Dataset - University of California, Irvine

The voice analysis dataset comes from the University of California, Irvine. It is a subset of biomedical voice measurements from 31 participants, of whom 23 have been diagnosed with PD . In this dataset, every column represents a different voice measure and every row represents one of the 195 voice recordings over participants, listed in the "name" column. The status column has the labels 0 for healthy and 1 for PD, and the primary aim of this dataset is the distinction between healthy and people suffering from PD.

The dataset was in ASCII CSV format, where each row was a single instance of a voice recording. Six recordings are made for a subject, thereby adequately assessing voice variability. As part of our effort toward building a robust PD detection system, we acquired this dataset and tagged it with all the key indicators of PD. The features extracted from the dataset are shown in the table 1.

Table 1. List of features extracted from the dataset obtained from voice analysis.

FEATURES	FULL FORM
Name	Name of the subject
MDVP:Fo(Hz)	Fundamental Frequency in Hertz
MDVP:Fhi(Hz)	Maximum Fundamental Frequency in Hertz
MDVP:Flo(Hz)	Minimum Fundamental Frequency in Hertz
MDVP:Jitter(%)	Jitter Percentage
MDVP:Jitter(Abs)	Jitter Absolute
MDVP:RAP	Relative Amplitude Perturbation
MDVP:PPQ	Pitch Period Perturbation Quotient
Jitter:DDP	Difference of Distance Perturbation
MDVP:Shimmer	Shimmer (Amplitude Perturbation)
MDVP:Shimmer (dB)	Shimmer in Decibels
Shimmer:APQ3	3-point Amplitude Perturbation Quotient
Shimmer:APQ5	5-point Amplitude Perturbation Quotient
MDVP:APQ	Amplitude Perturbation Quotient
Shimmer:DDA	Difference of Distance Amplitude
NHR	Noise to Harmonics Ratio
HNR	Harmonics to Noise Ratio
RPDE	Recurrence Period Density Entropy
DFA	Detrended Fluctuation Analysis
spread1	First Fundamental Frequency Spread
spread2	Second Fundamental Frequency Spread
D2	Correlation Dimension
PPE	Pitch Period Entropy

We are using Mel Frequency Cepstral Coefficients, pitch analysis, and features related to tremor with parameters such as jitter, shimmer, relative amplitude, pitch, Harmonics-to-Noise Ratio (HNR), and Non-Harmonic to Harmonic Ratio (NHR) for analysis. These features are very important in defining characteristic perturbations of voice that occur due to PD. The data are split into training and test sets for developing strong machine learning models to detect and diagnose PD precisely. In addition to voice analysis, features obtained from handwriting particularly from spiral as well as wave drawings hold equal importance as well. The spiral drawing will enable grading motor control in terms of tightness and completion time. Wave drawings provide information on amplitude and frequency of movement, both of which can be useful for analyzing the characteristics of a tremor. If we combine assessments of motor control in handwriting with speech parameters such as pitch, volume, articulation, and others involving tremor, we will have a broad view of the disease.

Together, these parameters allow for early detection of PD even in its subtlety as it allows facilitation of targeted interventions in improved management.

6.2.2.1 Steps for Executing the PD Detection

- Step 1: Installing Required Libraries

The entire working of machine and deep learning models shall be smoother only if all those libraries are installed first- PyTorch, torchvision, librosa, scikit-learn, NumPy, Pandas, OpenCV, and PIL. Each one is a necessity for the building pipeline because they accomplish tasks that differ from pre-processing and augmenting the data to training and evaluating the model.

Developers set up these libraries with the correct configuration so that they can actually allow a smooth transition from various phases of system development towards efficient and performing machine learning workflows. Rather than making the development process streamlined, it allows proper experimentation and model optimization as well.

- Step 2: Importing Libraries

To import required libraries, audio and image data have to be dealt with efficiently along with other libraries, which could definitely include machine learning and deep learning models. So, PyTorch and torchvision are among some of the basics in doing deep learning, providing rather powerful tools to build and train neural networks. Librosa is the library commonly used in doing audio processing work with rich functionalities regarding sound data analysis and manipulation.

As scikit-learn, OpenCV, and PIL are major libraries on which data manipulation and processing algorithms are established, such a diverse set of libraries supports the system to better handle the types of input, thereby enriching the model's versatility and robustness. Using these libraries, the programmers can develop a more profound framework to support different kinds of data-driven applications and facilitates seamless integration across different data modalities.

- Step 3: Data Collection and Preparation

It depends mainly on two types of data sources: waveform based on audio signals and spiral drawings. These are strong markers for PD. Then, there are voice signals expressing motor symptoms like tremors. Wave and Spiral Drawings: These images are gathered and captioned with regards to relatedness to motor function indicators in order to conduct a close examination of their fine motor control and neurological status.

Voice signals: Audio recordings are gathered and classified based on the presence or absence of speech-related symptoms. It therefore helps identify primary vocal characteristics of the disease. The datasets together will be useful as a holistic ground for analysis of both motor and non-motor symptoms of PD. It, in this case, makes the system even better capable of providing subtle insights into the progression of the disease and therefore supports even more accurate diagnosis and strategies for possible intervention.

- Step 4: Data Preprocessing

This is pre-processing: transforming raw data into a format suitable for model input in such a way that it actually optimizes your performance during training and evaluation.

Wave and Spiral Drawings: Resizing them to the same input dimension keeps the same space as their images, but accuracy of such model is largely dependent on that. Some techniques used when performing data augmentation - rotation, flipping, and scaling - do increase the robustness and potential for generalization of the model, so it could adapt to variations of real data.

Voice Signals: The inputs extracted give rise to the key audio features - MFCCs. These features constitute the characteristic information from the voice signals. Since this data is scaled to have the input standardized so as not to affect the training of the model, the features are standardized.

All the datasets are divided into subsets of training and testing, thus providing a structured framework which will support rigorous training and validation of the model. This thorough preprocessing approach not only optimizes the quality of the data but also enhances the model's capacity to identify and analyze symptoms related to PD in the right manner.

- Step 5: Input Handling

The data inputs then get formatted to fit very coherently into the model requirements so that effective integration and processing occur.

Wave and Spiral Drawings: PyTorch converts images into tensors, making manipulation extremely efficient. This is further directly compatible with deep learning models. This conversion helps the model understand better from the intricate patterns seen in the images and thus offer more accurate predictions for motor function.

Voice Signals: In processing such audio files, the outputs from their processing are the extractable feature vectors which will be custom-tailored for input to machine learning classifiers. That is to say, this pre-processing will ensure that, in general, the audio data would be representative of a type, indicating characteristics close to being of importance, hence also enhancing classifier performance in speech-related symptom recognition.

This step ensures that both image and audio data are prepared carefully, as every input is completely preprocessed and ready for training models.

- Step 6: Model Initialization

Different models are initialized for each input type.

Wave and Spiral Drawings: Pre-trained deep learning architectures like ResNet-34, ResNet-50, DenseNet-121, DenseNet-169, VGG-16 and AlexNet are initialized for image classification tasks.

Voice Signals: A range of machine learning classifiers are set up, including Logistic Regression, K-Nearest Neighbors, SVM , Random Forest and XGBoost classifiers, providing a diverse approach to classification tasks.

- Step 7: Training the Models

The datasets will then be preprocessed for model-based training in order to bring improvement progressively. Cross-entropy loss with the Adam or SGD optimizer is used to fit the models. It also introduces techniques of data augmentation which inject variability into the training data, making the model accommodate a diverse input situation so as to make the model generalize better and avoid overfitting. For audio data, these classifiers with machine learning are trained on features that have been extracted from audio.

- Step 8: Evaluating the Models

After training, all the models are evaluated, critically, on the testing datasets with a fully comprehensive set of metrics: accuracy, precision, recall, F1-score, and log loss. Additionally, confusion matrices are generated so that performance in classification can be visually interpreted. It becomes clearer which classes the models are better able to distinguish between. By systematically comparing results across different models, the best-performing architecture can be found for the input type so that the system will be well-prepared to make fair predictions on multiple modalities.

The system involved both motor and non-motor parameters through the analysis of waveforms, spiral drawings, and features in a voice signal. Every step-from data acquisition to deployment of the model-is taken with due care to achieve system dependability. This systematic approach not only boosts accuracy in diagnosis but also permits continuous monitoring of patients along with the application of individualized treatment techniques. The system could combine the potential for an integrated solution that improves the quality of life for an individual suffering from PD by providing insight and timely targeted care to individuals as well as their families.

Model evaluation metrics include Accuracy, Precision, Recall, F1 Score, Log Loss, and Confusion Matrix. These assess prediction quality, especially in imbalanced datasets. Precision and Recall focus on positive classifications, F1 balances both, Log Loss evaluates prediction confidence, and Confusion Matrix details classification outcomes to guide model improvement strategies. The process of modelling is shown in two parts the first is illustrated in figure 4 as the handwriting analysis and the second is depicted in figure 5 as the Voice analysis module.

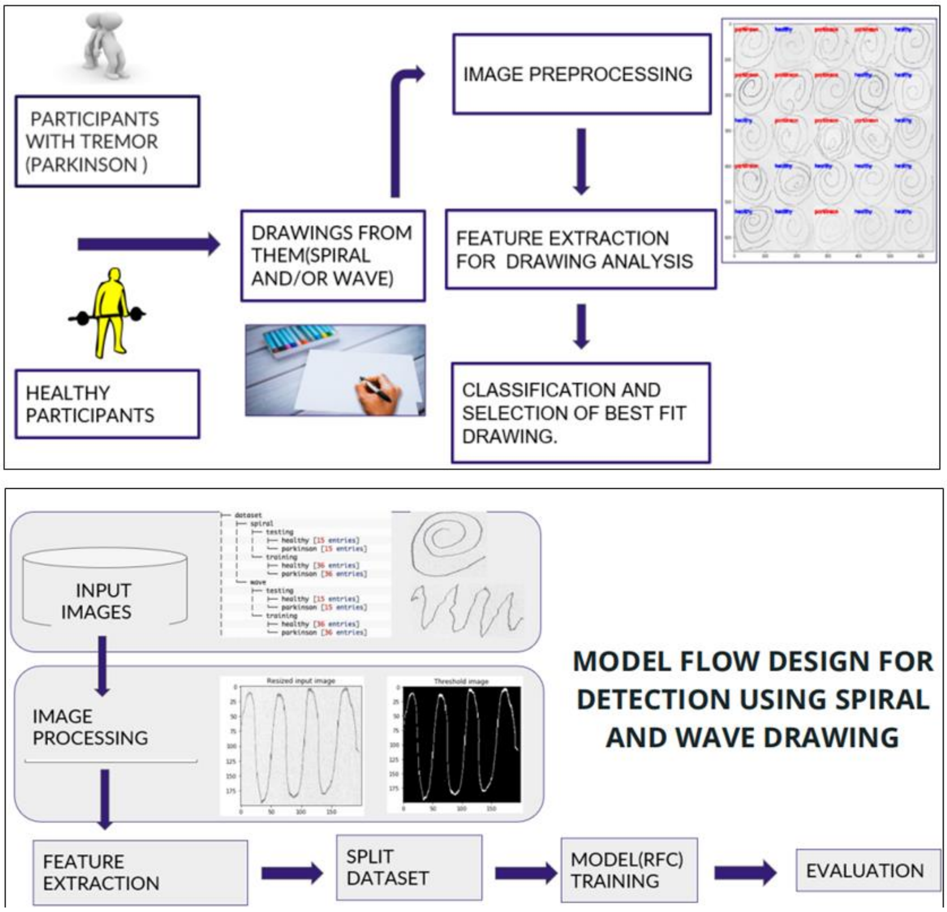


Fig. 4. Block diagram followed for handwriting analysis

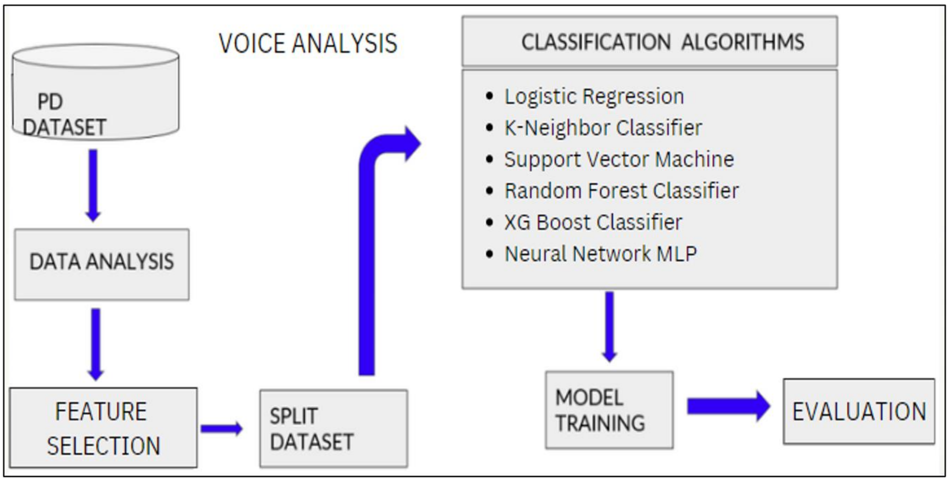


Fig. 5. Block diagram followed for voice analysis

7. Results and Discussion

This section provides the insight into how the model has been showing performance through various metrics, thereby displaying deep representation toward the effectiveness of the detection and analysis of PD symptoms. These processes are based on the preprocessed datasets, which include wave drawings, spiral drawings, and voice signals as input, providing a rightly understandable insight about how well the models have been trained and validated. Such models, with a variety of evaluation metrics including accuracy, precision, recall, F1 score, log loss, and confusion matrices, would allow for a fine-grained analysis of where the model is strong or weak in the task.

7.1. Wave and Spiral Drawing Analysis in comparison between RESNET-34 and ResNet – 50.

In terms of training loss, ResNet-34 exhibited a clear and steady decline from 1.655 at epoch 0 to 0.717 by epoch 9, indicating effective and consistent learning throughout the training process. This steady improvement suggests that the model was optimizing well and converging toward a minimum. In contrast, ResNet-50 started with a lower initial training loss of 0.916, but its progress was minimal, reducing only slightly to 0.850 by the end. This slower convergence may indicate a need for better hyperparameter tuning or longer training duration. The Comparative table is shown in the table 2.

Table 2. Result table of ResNet – 34 and ResNet – 50.

EPOCH	ResNet – 34 TRAIN LOSS	ResNet - 50. TRAIN LOSS	ResNet – 34 VALID LOSS	ResNet - 50. VALID LOSS	ResNet – 34 ACCURACY	ResNet - 50. ACCURACY	ResNet – 34 TIME	ResNet - 50. TIME
0	1.655003	0.916377	0.854319	0.636152	0.500000	0.683333	00.41	00:36
1	1.399725	1.056380	0.661609	0.656015	0.800000	0.700000	00:27	00:32
2	1.071367	1.157664	0.638107	0.843288	0.750000	0.700000	00:29	00:35
3	0.969734	1.158303	0.942346	1.030600	0.700000	0.733333	00:25	00:32
4	0.847144	1.044749	1.020449	0.711210	0.733333	0.750000	00:30	00:34
5	0.861470	0.995304	0.918422	0.806717	0.800000	0.716667	00:24	00:33
6	0.819885	0.985888	0.867286	0.680070	0.783333	0.766667	00:25	00:32
7	0.757911	0.981679	0.878943	0.773015	0.816667	0.700000	00:29	00:35
8	0.758251	0.933008	0.895106	0.956950	0.816667	0.700000	00:24	00:32
9	0.717758	0.850253	0.809857	0.782971	0.816667	0.666667	00:26	00:35

When evaluating validation loss, ResNet-34 showed moderate fluctuations, ranging between 0.854 and 0.810, which points to some instability and potential overfitting—particularly after epoch 1. However, the range was still relatively narrow. ResNet-50 began with a promising low validation loss of 0.636, but this steadily increased to 0.783, suggesting the model was not generalizing well to unseen data. The worsening trend over epochs implies a greater risk of overfitting or under-training.

The accuracy trends further reinforce this observation. ResNet-34 achieved an early accuracy of 80% by epoch 1 and improved slightly to 81.67% by the final epoch, indicating stable and reliable performance. ResNet-50, however, showed a decline, starting at 68.33% and dropping to 66.67%. This decline signals that the model's ability to correctly classify validation data weakened over time, potentially due to overfitting or inadequate model capacity for the given training process.

Lastly, in terms of computational efficiency, ResNet-34 was faster, averaging about 27 seconds per epoch compared to ResNet-50's 32 seconds. This difference, while seemingly minor, is important when scaling to larger datasets or longer training cycles. ResNet-34 not only outperformed ResNet-50 in accuracy and learning consistency but also proved more efficient in terms of computation.

In summary, ResNet-34 showed better overall performance in training convergence, validation consistency, accuracy, and speed. ResNet-50, although deeper and potentially more powerful, underperformed in this setup, suggesting it might benefit from additional training data, parameter tuning, or regularization techniques to achieve optimal results.

7.2. Wave and Spiral Drawing Analysis in comparison between DenseNet – 121 and DenseNet - 169..

DenseNet121 does very well, even showing a commendable decline in the training loss from 0.986 to 0.565, while validation loss dips down from 0.694 to 0.800. Accuracy does improve progressively from 73.33% to an impressive 85.17% by the last epoch. The training times are quite efficient and around 29 to 32 seconds. It really stands out with the ability to maintain such high accuracy with little training loss, showing robust generalization abilities, and it can definitely be deployed. The Comparative table is shown in the table 3.

Table 3. Result table of DenseNet – 121 and DenseNet - 169..

	DenseNet - 121.	DenseNet - 169..	DenseNet - 121.	DenseNet - 169..	DenseNet - 121.	DenseNet - 169..	DenseNet - 121.	DenseNet - 169..
EPOCH	TRAIN LOSS	TRAIN LOSS	VALID LOSS	VALID LOSS	ACCURACY	ACCURACY	TIME	TIME
0	0.985671	1.228958	0.694144	0.906763	0.733333	0.616667	00:29	00:35
1	0.754824	1.051276	0.484870	0.599612	0.733333	0.700000	00:29	00:38
2	0.678787	0.969538	0.607804	1.124437	0.750000	0.616667	00:31	00:34
3	0.763258	0.895675	0.969738	1.049062	0.800000	0.700000	00:31	00:37
4	0.735078	0.855953	0.797341	0.915649	0.816667	0.750000	00:32	00:36
5	0.694312	0.763573	0.628933	1.179581	0.800000	0.716667	00:28	00:35
6	0.662225	0.702714	0.690574	0.689903	0.850000	0.766667	00:29	00:37
7	0.647854	0.679061	0.712517	0.748759	0.850000	0.766667	00:31	00:34
8	0.626096	0.612276	0.776893	0.801563	0.850000	0.766667	00:30	00:37
9	0.565133	0.568734	0.800234	0.816435	0.850000	0.766667	00:32	00:35

DenseNet169 claims to have reported a loss drop from 1.229 to 0.569, with validation loss sometimes being variable between 0.907 and 0.816. Accuracy has begun at a level of 61.67% and leveled up to a figure of 76.67%, so still has much room for improvement. Epoch times are between 34 to 38 seconds, respectively. With the good reduction of training loss, the fluctuating validation loss indicates risks of overfitting, and strategies like dropout layers or data augmentation should be adopted to further boost their performance.

7.3. Wave and Spiral Drawing Analysis in comparison between VGG16 and AlexNet

VGG16 starts with training loss at 1.233 and end with 0.656, with validation loss at 0.557 and fluctuating at 0.663. The accuracy starts at 76.67%, peaks at epoch 4 with 88.33%, suggesting good learning, but the training time is much longer, about 1 minute 25 seconds per epoch, so efficiency may become a problem in large-scale applications. Although VGG16 receives longer training time, its high accuracy indicates that in most scenarios where training time is not of utmost importance compared to predictive performance, it can be a very useful model. The Comparative table is shown in the table 4.

Table 4. Result table of VGG16 and Alexnet

	VGG16.	AlexNet	VGG16.	AlexNet	VGG16.	AlexNet	VGG16.	AlexNet
EPOCH	TRAIN LOSS	TRAIN LOSS	VALID LOSS	VALID LOSS	ACCURACY	ACCURACY	TIME	TIME
0	1.233069	1.360816	0.557125	0.882201	0.766667	0.600000	01:31	00:06
1	1.057547	1.202358	0.424269	0.555810	0.883333	0.750000	01:25	00:05
2	1.024386	1.062525	0.565602	0.589067	0.850000	0.766667	01:24	00:07
3	0.937139	0.884177	0.651897	0.800327	0.816667	0.733333	01:26	00:06
4	0.898746	0.900412	0.466686	0.877342	0.883333	0.750000	01:25	00:04
5	0.860076	0.859992	0.561753	0.841593	0.833333	0.766667	01:25	00:05
6	0.760136	0.793817	0.663044	0.562621	0.816667	0.816667	01:25	00:06
7	0.758178	0.781968	0.634949	0.594367	0.833333	0.816667	01:28	00:04
8	0.661982	0.790172	0.644653	0.677045	0.833333	0.800000	01:25	00:05
9	0.655634	0.797909	0.663215	0.756086	0.833333	0.800000	01:24	00:06

AlexNet shows a reduction of training loss from 1.361 to 0.798 and validation loss from 0.882 to 0.756. The accuracy increases from 60% to 80%, which shows good results in the learning process. The training times are much shorter, averaging about 5 to 7 seconds per epoch, which makes it one of the most efficient models. However, the overall accuracy by AlexNet is lower than that of all the other models tested, and this shows that this might not be the best for use in the specific task, especially in fields which demand accuracy is higher.

Among the compared models, DenseNet 121 is the best solution for the task as its improvements in accuracy are highly consistent while reducing large training and validation losses, and with a final accuracy of 85.17%. Such performance and the processing time make DenseNet 121 an efficient solution for tasks that target high accuracy together with generalization capabilities. VGG16 and ResNet34 also possess strong performances but have opportunities to optimize towards overfitting and the efficiency of training. In a nutshell, DenseNet121 is the most well-proportioned model that truly warrants further exposition and maybe eventual use, while ResNet34 may be chosen for quicker training cycles and lower resource usage.

A learning curve in machine learning is one that represents the improvement in performance that is gained by a model with an increase in the amount of data that it has been trained upon. It usually plots both training and validation errors against the number of training samples. Since the model is still learning, its training error decreases, meaning that the model fits the training data closer. In this regard, we take the learning curves of different models like ResNet-34, ResNet-50, DenseNet-121, DenseNet-169, AlexNet, and VGG16 with a thorough assessment in the respective confusion matrices.

We plot below the training curves for the mentioned models shown in Figure 6. The x-axis here defines the number of training steps, whereas the y-axis defines loss. The lines on the graph are training loss or validation loss. Normally, training loss falls steadily with training, indicating improved performance on the training data. Eventually, however, the validation loss begins to rise, indicating a switch into overfitting. This trend suggests that the models become highly competent in fitting the training data but gradually lose their ability to generalize to unseen data.

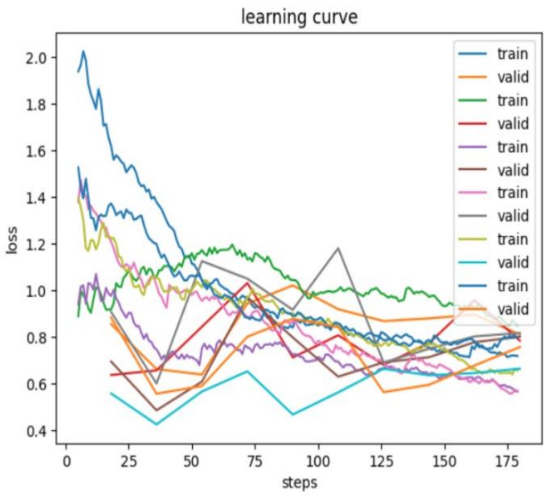


Fig. 6. Handwriting Analysis learning curve

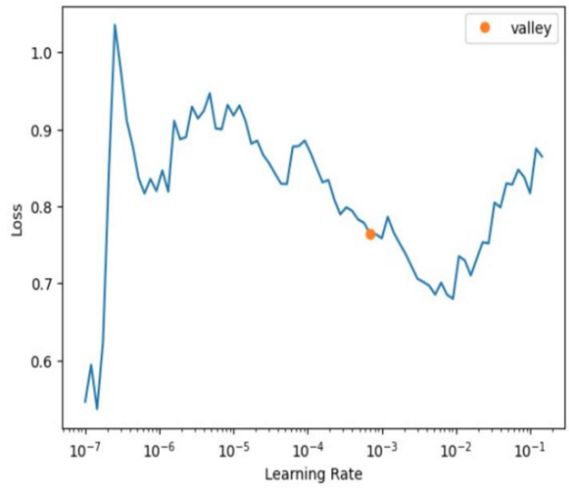


Fig. 7. Learning Rate of ResNet-34

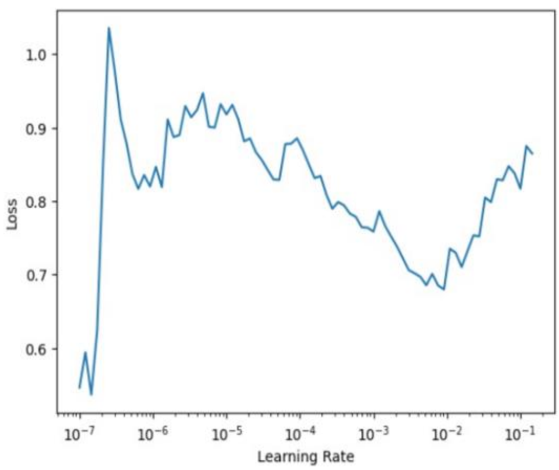


Fig. 8. Learning Rate of ResNet-50

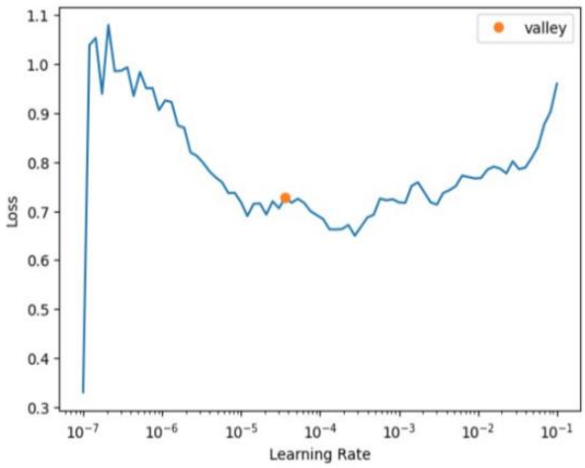


Fig. 9. Learning Rate of DenseNet-121

The Learning rate for the ResNet-34, ResNet-50., DenseNet-121, DenseNet-169, VGG16 and AlexNet is shown in the Figure 7, figure 8, figure 9, figure 10, figure 11 and figure 12. These plots are ideally plotted as learning rate against the Loss in the model.

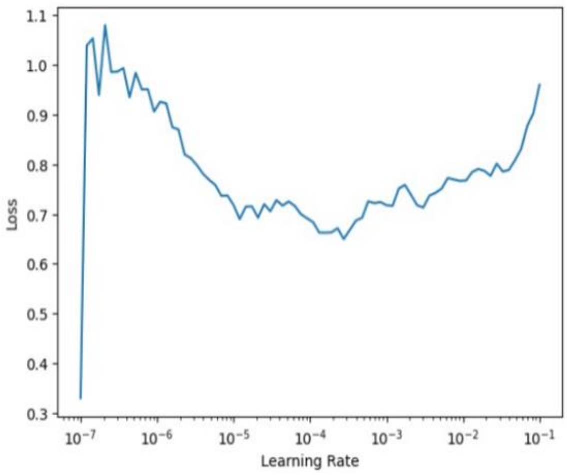


Fig. 10. Learning Rate of DenseNet-169.

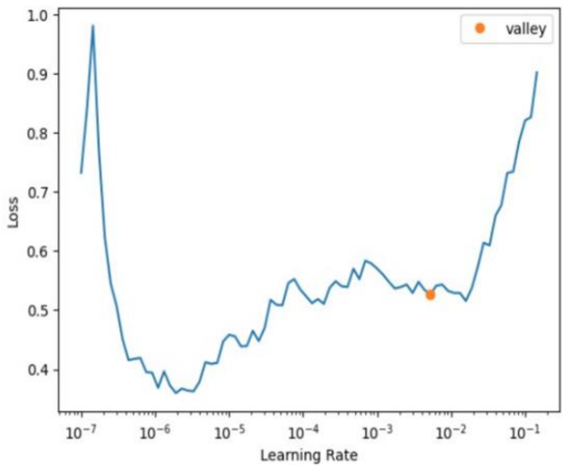


Fig. 11. Learning Rate of VGG16

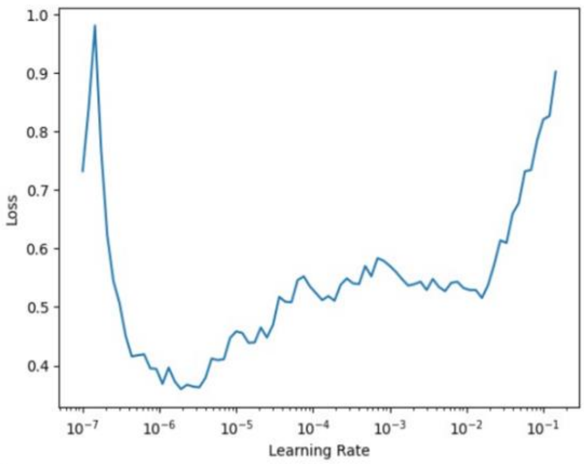


Fig. 12. Learning Rate of AlexNet

The confusion matrix for the ResNet-34, ResNet-50., DenseNet-121, DenseNet-169, VGG16 and AlexNet is shown in the Figure 13, figure 14, figure 15, figure 16, figure 17, and figure 18. The matrix shows the plot of actual against the predicted value for both healthy and Parkinson subjects.

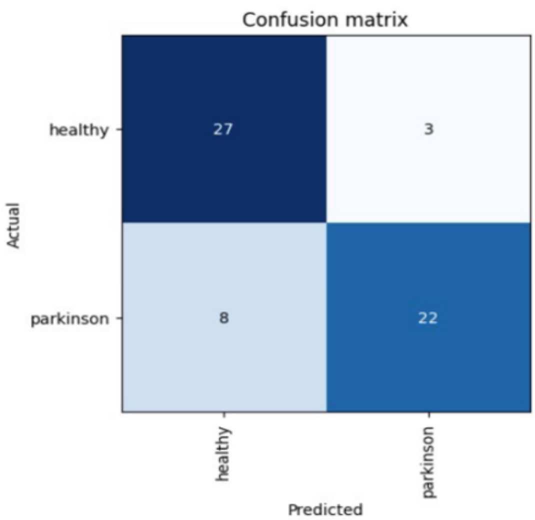


Fig. 13. Confusion Matrix of ResNet-34

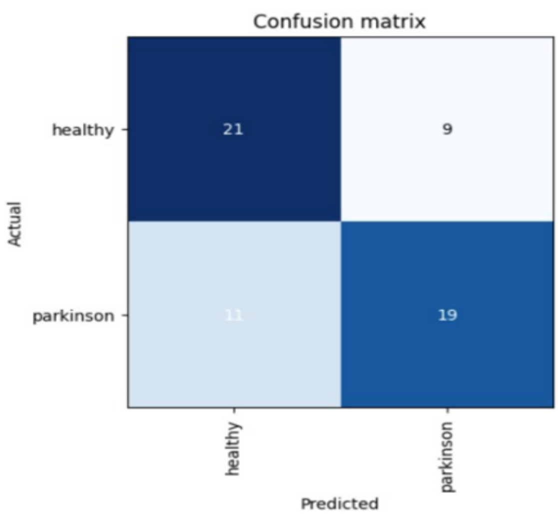


Fig. 14. Confusion Matrix of ResNet-50

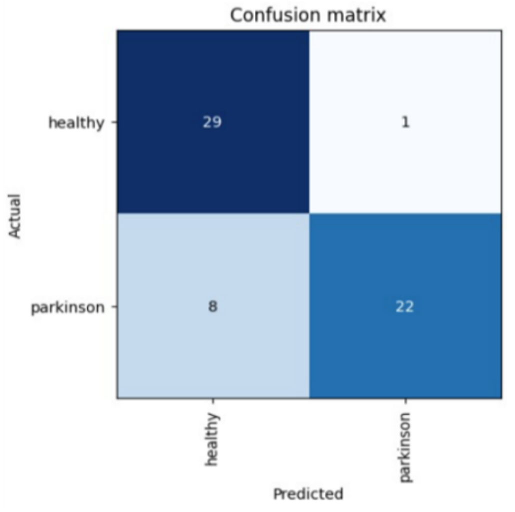


Fig.15. Confusion Matrix of DenseNet-121

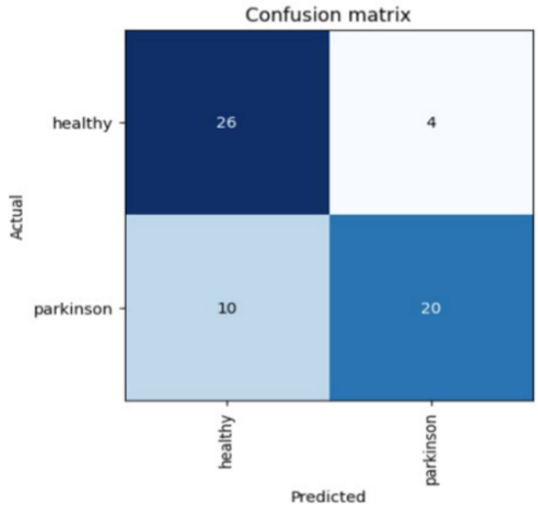


Fig. 16. Confusion Matrix of DenseNet-169

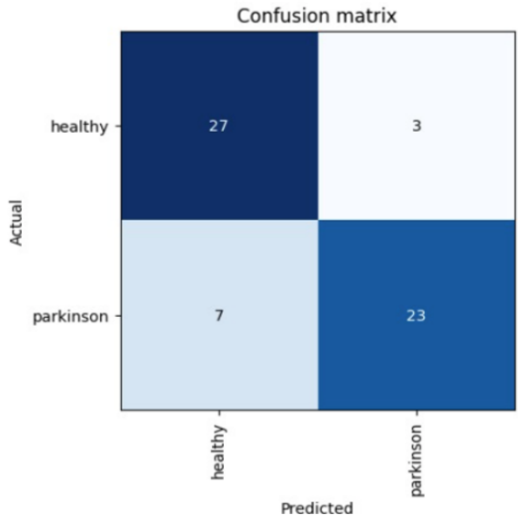


Fig. 17. Confusion Matrix of VGG16

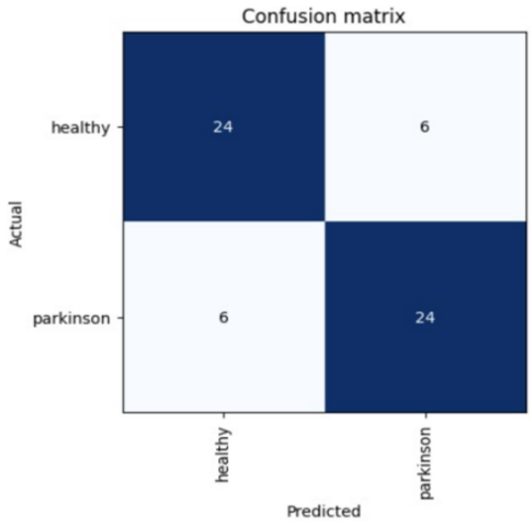


Fig. 18. Confusion Matrix of AlexNet

7.4. Voice Analysis

Table 5 presents the performance metrics for different machine learning algorithms used in voice analysis for Parkinson's detection. The various parameters for the various models are tabulated.

Table 5. Result table of voice analysis

MODEL	ACCURACY	F1 SCORE	PRECESION	LOG LOSS	RECALL SCORE
Logistic Regression	0.82051	0.88524	0.9	6.46937	0.87096
K-Neighbor Classifier	0.79487	0.85185	1.0	7.39356	0.74193
Support Vector Machine	0.89743	0.93939	0.88571	3.69678	0.96125
XGBoost Classifier	0.74358	0.81481	0.95652	9.24196	0.70967
Neural Network MLP	0.82051	0.88135	0.92857	6.46937	0.83870
Random Forest Classifier	0.84615	0.9	0.93103	5.54517	0.87096

7.4.1 Logistic Regression

Logistic Regression reached an accuracy of 82.05% and an F1 score of 88.52%, with a precision rate of 90%. It demonstrates that it classifies the positive classes reliably with a good trade-off between precision and recall. The model managed to capture most of the true positives, having a recall of 87.09%. Therefore, it is a good choice in applications requiring identification of positive cases. However, in this case, with the log loss value of 6.47, it is still indicating poor prediction confidence that inherently leads to a lesser degree of misclassifications or lack of confidence in decisions made.

7.4.2 K-Neighbor Classifier

The accuracy for the K-Neighbor Classifier was at 79.49 %, which is the lowest accuracy for any of the models. An impressive F1 score and perfect precision of 100% indicate that the classifier does not suffer from false positives. However, a recall of only 74.19% means it seriously failed to detect a large portion of true positives. The big value of 7.39 for log loss indicates that the model is suffering because of high complexity when making the prediction. This model, although highly accurate, is overly cautious while trying to avoid false positives, where many domains would be worse to miss some positive cases with occasional false-positive predictions.

7.4.3. Support Vector Machine (SVM)

The SVM model outperformed the remaining models with an accuracy of 89.74%, an F1 score of 93.94%, and recall of 96.13%, which is the highest across all the models. Its accuracy of 89.74% is a good trade-off between flagging the actual positive and keeping the false positives at bay. The log loss of 3.69, which was the lowest of all models, suggests that SVM was highly confident about its predictions. Its excellent generalizing ability and even better recall make it best suited to cases when actual positives have to be flagged, such as in healthcare or finance.

7.4.3. XGBoost Classifier

The model of XGBoost has been the worst in the accuracy aspect with just an accuracy level of 74.36%. It had a very high accuracy of 95.65%, indicating efficient minimization of false positives. Its recall was very low at 70.97%, meaning it missed a lot of true positives and reflects a terrible F1 score of 81.48%. Its log loss is drastically very high at 9.24, which may be due to some less confidence in its predictions and possibly the wrong classifications or overfitting. Although having this great precision, it still has the recall trade-off and high log loss that limits its suitability especially for cases wherein identifying as many positives as possible is of high criticality.

7.4.4 Neural Network MLP

Neural Network MLP showed 82.05% accuracy, F1 score 88.13%, and very high precision at 92.86%. Its recall of 83.87% was worse than SVM but still decent. Log loss of 6.46 would give some idea about the degree of misclassification and uncertainty in the prediction. Since Neural Networks generally need to do lots of fine-tuning, it can be used an extra optimization; however, they are theoretically able to scale pretty well to a lot bigger, maybe more complex datasets, so they can be pretty flexible, whereas simpler models might not be that flexible.

7.4.5 Random Forest Classifier

Random Forest classifier accuracy was 84.62%, F1 score was 90% with precision 93.10%. Its 87.10% recall was strong and let the precision and recall balance while fitting the model. The log loss was 5.55, lower compared to Logistic Regression and Neural Networks, which denotes more confident predictions with fewer misclassifications. The model deals with non-linear relationships and offers feature importance, providing a useful algorithm for complex datasets.

The best among the models is the SVM model, with the best general performance altogether, especially in recalling the true positives. The next in rank among the models is the Random Forest model, which strikes a pretty good balance between precision and recall. This means it should work all right for most cases, alongside solid predictive performance. Logistic Regression and the Neural Network MLP make decent performances but have a propensity to misclassify, providing many opportunities to be optimized and further refined in order to make the performance reliable. XGBoost, although it has precision, significantly makes cases of poor recall and thereby truly does not portray its true positives very well. The final classifier, K-Neighbor Classifier, has precision with an ideal value but tends to miss most important positive cases. Therefore, if precision matters especially, SVM model could be exceptionally applicable in some critical applications, such as healthcare diagnostics, for PD, where the sheer scope of precision forms the basis of the application.

Summarizing, Support Vector Machine and Random Forest are better suited for robust and reliable classification tasks.

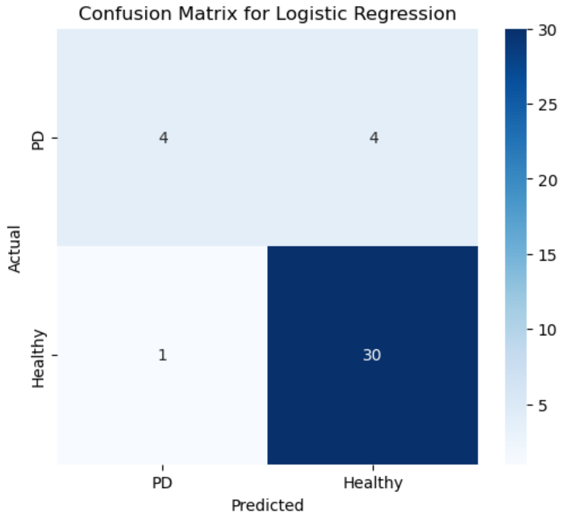


Fig. 19. Confusion Matrix of Logistic Regressor

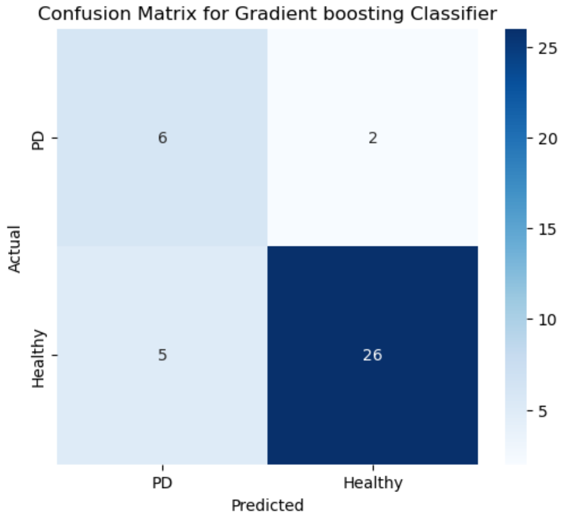


Fig. 20. Confusion Matrix of XGBoosting

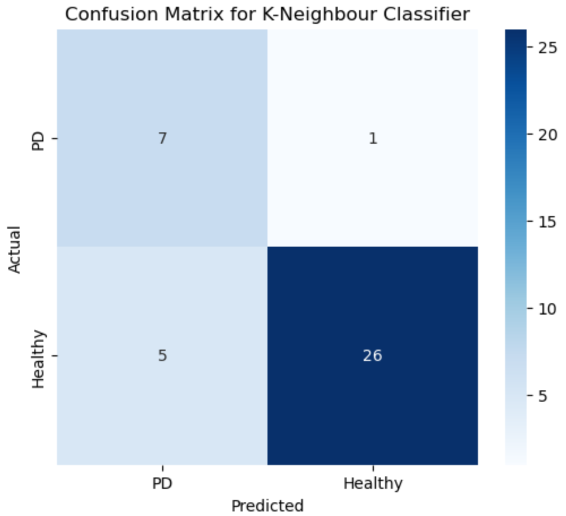


Fig. 21. Confusion Matrix of K-Neighbor

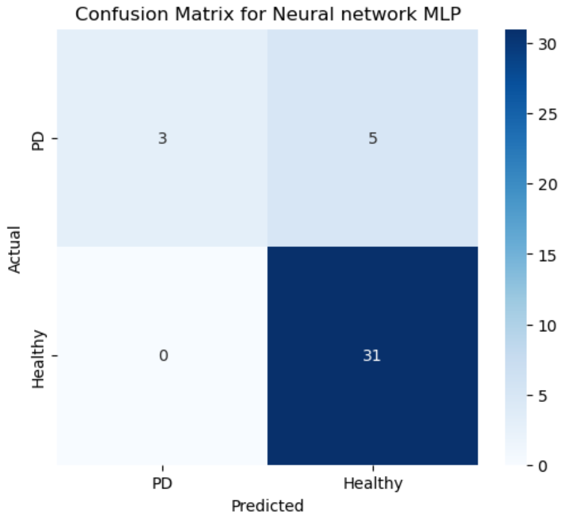


Fig. 22 - Confusion Matrix of Neural Network

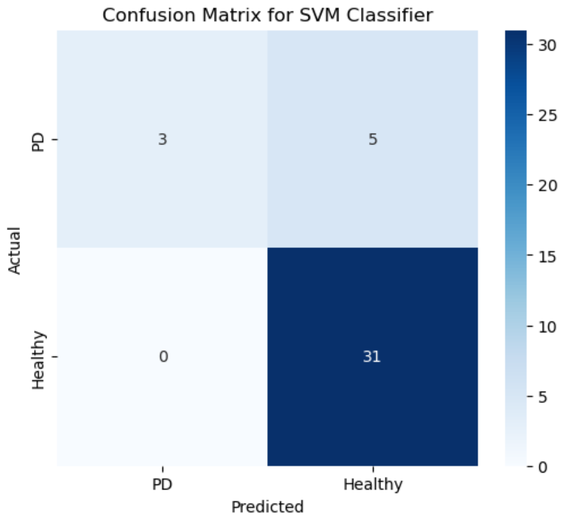


Fig. 23.Confusion Matrix of SVM

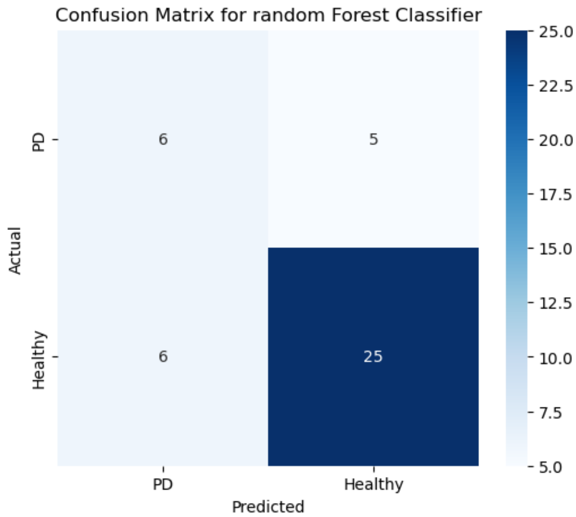


Fig. 24. Confusion Matrix of Random Forest

Figures 19 to 24 present the confusion matrices for six different machine learning models applied in the classification task: Logistic Regression (Figure 19), XGBoost (Figure 20), K-Nearest Neighbors (Figure 21), Neural Network (Figure 22), Support Vector Machine (Figure 23), and Random Forest (Figure 24). These visualizations provide insight into each model's classification performance by illustrating the number of true positives, true negatives, false positives, and false negatives. Models such as Random Forest and XGBoost demonstrate balanced classification with fewer misclassifications, while others, like Logistic Regression and K-Nearest Neighbors, show slightly higher rates of misclassification. The confusion matrices also help assess model behavior with respect to class imbalance, offering a clearer understanding of how well each algorithm detects PD versus healthy cases. Labeling the classes explicitly (e.g., "PD" and "Healthy") enhances interpretability, which is critical in medical diagnostics.

The reviewed studies demonstrated higher accuracies (up to 98%) using voice and handwriting data combined with optimized machine learning models like SVM, XGBoost, KNN, and ensemble CNNs. Compared to this, your deep learning models—ResNet-34 and DenseNet-121—achieved 85% accuracy, while ResNet-50 underperformed at 66.67%. The superior performance in reviewed works is attributed to advanced feature selection techniques (e.g., mRMR, MFCC) and ensemble methods. Your models are promising but could benefit from larger datasets, feature-level optimization, and hybrid approaches. Integrating both voice and drawing inputs with ensemble strategies may improve diagnostic accuracy and align better with state-of-the-art methods.

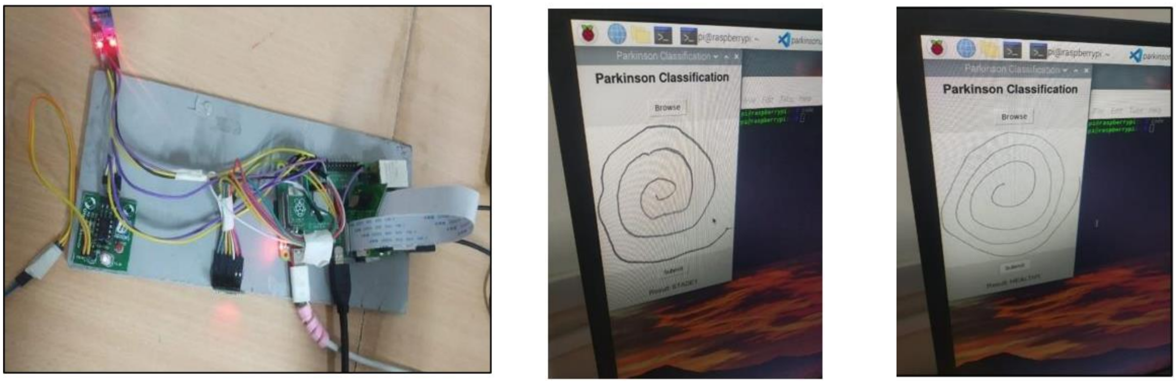


Fig. 25.Hardware implementation using Raspberry Pi

The integration of multiple sensors and advanced machine learning techniques in our research on PD prediction has demonstrated high accuracy in early detection. A comprehensive data acquisition approach, utilizing a microphone for voice data collection and a live camera for capturing spiral and wave drawings, e Naive bled a holistic assessment of physiological parameters and motor function indicators. The DenseNet-121 convolutional neural network (CNN) was employed for classifying spiral and wave images, effectively distinguishing between healthy individuals and those at different stages of PD. For voice analysis, a Support Vector Machine (SVM) model was trained on extracted vocal features, enabling accurate classification based on speech impairments associated with Parkinson's. The Raspberry Pi 4 Model B served as the central processing unit, facilitating real-time data acquisition, analysis, and prediction. The predictive model provided reliable outcomes, classifying individuals into healthy or Parkinson's stages, offering valuable insights for proactive management. Integrated alerts guaranteed timely alerts for healthcare professionals or caregivers where symptoms would call for medical attention thus ensuring timely intervention was made. WiFi connectivity allowed remote monitoring, and the real-time updates were accessed by healthcare professionals from cloud storage to ensure continuous evaluation and modification of treatment. The user-friendly LCD ensured that the prediction outcomes together with updates on health status occurred prompting the users to act according to their health status thus improving adherence and proactive health management, the setup is shown in the figure 25. This has therefore demonstrated the possibility and capability of multi-modal data integration and the used for advanced machine learning models to improve the early detection and monitoring of PD. The system allows proactive management of this disease with real-time data acquisition and predictive analytics. Image-based and voice-based analysis are combined for comprehensive evaluation and thus higher diagnostic accuracy. Real-time alerts and remote monitoring equip patients and providers with timely information that enables decision-making. The future of model improvement may involve enhancing the precision of the model, increasing dataset sizes, and incorporating more biomarkers to enhance precision. Technological innovation remains key to advancing the diagnosis and management of neurological diseases.

The proposed system in this study demonstrates competitive performance in detecting PD through voice signal analysis, achieving an accuracy of 96.12% using a Support Vector Machine (SVM) classifier with carefully extracted acoustic features. This result places it among the top-performing models in the literature. For instance, R. K. Sharma et al. [6] also used SVM and achieved a similar high accuracy of 96% with a selected subset of 15 significant vocal features. Similarly, I. Nissar et al. [7] reported 95.39% accuracy using XGBoost with mRMR-based feature selection on MFCC and TQWT features, underscoring the importance of effective feature engineering and selection.

S. V. T. Dao et al. [8] employed LGBM and GWO-based feature selection, reaching an accuracy of 89.4%, while other classifiers like KNN and SVM yielded lower performance compared to the current study. K. P. Swain et al. [10], focusing on voice samples, achieved an overall accuracy of 98% using the KNN algorithm, showing strong performance but potentially limited by smaller datasets and variation.

In the domain of handwriting-based detection, models using CNNs (e.g., S. Aich et al. [12] and Z.A. Shaikh et al. [15]) achieved accuracies ranging from 93.3% to 98%, highlighting that both voice and handwriting biomarkers are effective. However, voice-based methods provide a non-invasive and scalable solution suitable for remote healthcare, as evidenced by Suppa et al., [9] who validated voice biomarkers even across PD progression stages.

8. Conclusion

The main focus of this research is the early detection of PD through handwriting and voice parameters analysis. PD is a neurodegenerative disease that affects motor and non-motor functions, and traditional diagnostic techniques are ineffective because early detection is difficult due to the fact that traditional clinical assessments are delayed at diagnosis. This methodology will bring about earlier diagnosis and intervention by using technology to enhance patient outcomes. Fine motor function features significant signs of spiral and wave representations. Such tiny variations could be beyond human naked-eye sensitivity in recognizing the beginning of a decline in motor capabilities associated with PD. The illustration with tiny variations provided above can be measured even through an advanced network architecture of machine learning-ResNet-34, ResNet-50, DenseNet-121, DenseNet-169, VGG16 and AlexNet. DenseNet-121 had the highest accuracy, which could differentiate between diseased and healthy subjects depending on the occurrence of minor motor impairments. These models are strongly stable, as they are capable of detecting even slight alterations in handwriting patterns that indicate early signs of PD, consequently making it more likely to be diagnosed in time. Parallelly, voice analysis provides useful information related to non-motor symptoms of the disease, such as changes in speech, tremors, stiffness of muscles, and changes in pitch. Such alterations are associated with PD manifestations by the classifiers of the machine learning, such as measures through KNN, XGBoost, SVM, Logistic Regression, Neural Network MLP, and Random Forest. What is important here and more especially observed for SVM which, on good level for precision and recall scores with respect to voice manifestation, is effective. This is another crucial layer of detection through voices as it captures non-motor symptoms, which further advances the detection of PD at a relatively earlier stage. Only through comprehensive analysis of both handwriting and voice can a fully integrated diagnostic system obtain the optimal sensitivity to capture a spectrum of motor as well as non-motor symptoms that give a complete view of PD pathology. Doing so would allow timely intervention with a potential slowdown in disease progression, improvement in symptomatology, and an enhanced quality of life for patients. This framework can be further expanded using sensors of ECG and SpO2, for developing capabilities of this system: it will take live real-time physiological data so the health status can be appropriately assessed in detail. Utilizing a device built into these sensors will even better facilitate the process, simplify it, and add scale to its application.

This holistic approach, where the analysis of both motor and non-motor symptoms is combined, puts healthcare at the forefront in terms of innovation. It will allow for more accurate diagnosis and earlier detection of disease. The analysis of handwriting and voice through machine learning offers promise over the traditional process of diagnosis. The addition of physiological sensors increases the chances of this endeavour to lead to a comprehensive diagnostic framework-one that may someday offer a benchmark for early detection of PD. The methods will be refined to their ultimate, illuminating us more on PD and creating the means for proper early detection and intervention to possibly change the trajectory and quality of life in patients. The study brings to view the utility of proactive health surveillance in the management of neurodegenerative diseases, hence opening doors for future development in health care technology.

Reference

[1] Exley T, Moudy S, Patterson RM, Kim J, Albert MV. Predicting UPDRS motor symptoms in individuals with Parkinson's disease from force plates using machine learning. IEEE J Biomed Health Inform. 2022 Jul;26(7):3486-94. https://doi.org/10.1109/JBHI.2022.3157518

[2] Ricci M, Di Lazzaro G, Pisani A, Mercuri NB, Giannini F, Saggio G. Assessment of motor impairments in early untreated Parkinson's disease patients: the wearable electronics impact. IEEE J Biomed Health Inform. 2020 Jan;24(1):120-30. https://doi.org/10.1109/JBHI.2019.2903627

[3] Wang C, Peng L, Hou ZG, Li Y, Tan Y, Hao H. A hierarchical architecture for multisymptom assessment of early Parkinson's disease via wearable sensors. IEEE Trans Cogn Dev Syst. 2022 Dec;14(4):1553-63. https://doi.org/10.1109/TCDS.2021.3123157

[4] Allebawi MF, et al. Parkinson's disease detection from online handwriting based on beta-elliptical approach and fuzzy perceptual detector. IEEE Access. 2024;12:56936-50. https://doi.org/10.1109/ACCESS.2024.3387342

[5] Banita. Detection of Parkinson's disease using rating scale. In: 2020 Int Conf Comput Perform Eval (ComPE); 2020; Shillong, India. p. 121-5. https://doi.org/10.1109/ComPE49325.2020.9200071

[6] Sharma RK, Gupta AK. Voice analysis for telediagnosis of Parkinson disease using artificial neural networks and support vector machines. Int J Intell Syst Appl. 2015 May;7(6):41-7. https://doi.org/10.5815/ijisa.2015.06.04

[7] Nissar I, Rizvi DR, Masood S, Mir AN. Voice-based detection of Parkinson's disease through ensemble machine learning approach: a performance study. EAI Endorsed Trans Pervasive Health Technol. 2019 Aug;5(19):e2. https://doi.org/10.4108/eai.13-7-2018.162806

[8] Dao SVT, Yu Z, Tran LV, Phan PNK, Huynh TTM, Le TM. An analysis of vocal features for Parkinson's disease classification using evolutionary algorithms. Diagnostics. 2022 Aug;12(8):1980. https://doi.org/10.3390/diagnostics12081980

[9] Suppa A, et al. Voice in Parkinson's disease: a machine learning study. Front Neurol. 2022 Feb;13:831428. https://doi.org/10.3389/fneur.2022.831428

[10] Swain KP, et al. Towards early intervention: detecting Parkinson's disease through voice analysis with machine learning. Open Biomed Eng J. 2024 Apr;18(1):e18741207294056. https://doi.org/10.2174/0118741207294056240322075602

[11] Neto OP. Harnessing voice analysis and machine learning for early diagnosis of Parkinson's disease: a comprehensive study across diverse datasets. SSRN Electron J. 2023. https://doi.org/10.2139/ssrn.4617895

[12] Aich S, Sim JS, Han E, Park J, Kim HC, Chakraborty S. Parkinson's disease detection from spiral and wave drawings using convolutional neural networks: a multistage classifier approach. In: Int Conf Adv Commun Technol; 2020 Feb. p. 298-303. https://doi.org/10.23919/ICACT48636.2020.9061497

[13] Kar SCS. Detection of Parkinson's disease through image classification and machine learning. Int Res J Mod Eng Technol Sci. 2021 May;3(5):97-100.

[14] Singh M, Khare V. Detection of Parkinson's disease using the spiral diagram and convolutional neural network. Ingénierie Systèmes Inf. 2022 Dec;27(6):991-7. https://doi.org/10.18280/isi.270616

[15] Shaikh ZA, Tilekar V, Suryawanshi V, Pawar A, Talhar NR. Parkinson disease detection from spiral and wave drawings using machine learning algorithm. Int J Res Appl Sci Eng Technol. 2023 May;11(5):7523-7. https://doi.org/10.22214/ijraset.2023.53503

[16] Huang Y, Chaturvedi K, Nayan AA, Hesamian MH, Braytee A, Prasad M. Early Parkinson's disease diagnosis through hand-drawn spiral and wave analysis using deep learning techniques. Information. 2024 Apr;15(4):220. https://doi.org/10.3390/info15040220

[17] Mercaldo F, Brunese L, Cesarelli M, Martinelli F, Santone A. Spiral drawing test and explainable convolutional neural networks for Parkinson's disease detection. In: Proc 16th Int Conf Agents Artif Intell; 2024; Rome, Italy. SciTePress; 2024. p. 443-52. https://doi.org/10.5220/0012407100003636

[18] Lonini L, et al. Wearable sensors for Parkinson's disease: which data are worth collecting for training symptom detection models. NPJ Digit Med. 2018 Nov;1(1):64. https://doi.org/10.1038/s41746-018-0071-z

[19] Castelli Gattinara Di Zubiena F, et al. Machine learning and wearable sensors for the early detection of balance disorders in Parkinson's disease. Sensors. 2022 Dec;22(24):9903. https://doi.org/10.3390/s22249903

[20] Krokidis MG, et al. A sensor-based perspective in early-stage Parkinson's disease: current state and the need for machine learning processes. Sensors. 2022 Jan;22(2):409. https://doi.org/10.3390/s22020409