International Journal of Innovative Research and Scientific Studies, 8(3) 2025, pages: 2906-2916

ISSN: 2617-6548

URL: www.ijirss.com

Optimized technique for the early identification of Parkinson's disease using machine learning-based handwriting and voice analysis

Nagarathna C R[1], Jobby George[2], Komala H K[3], Kavitha M[4], K R Swetha[5*]

[1]*BNM Institute of Engineering, Bangalore, Karnataka, India.*
[2]*College of Health Sciences, University of Fujairah, UAE.*
[3]*Community Health Nursing, BGS Institute of Nursing Sciences, Mandya, Karnataka, India.*
[4]*JSS Academy of Technical Education, Bengaluru, Karnataka, India.*
[5]*Computer Science and Engineering, BGS Institute of Technology, Adichunchanagiri University, Mandya, Karnataka, India.*

Corresponding author: K R Swetha (*Email: swethagowdha@gmail.com*)

Abstract

Movement impairments caused by Parkinson's disease include tremors and rigidity in the muscles. Due to the fact that each patient's symptoms are unique, PD can easily go undiagnosed. Because of this, the early phases of Parkinson's disease, when symptoms are modest, are not well diagnosed with existing procedures. New research, however, supports the idea that PD may be identified early by using handwriting and speech abnormalities as indicators. Early detection of Parkinson's disease is essential since it guarantees that the illness's initial occurrence is identified and treated as soon as feasible. In addition, this method allows doctors to better control how the illness progresses. In the end, patients are able to prolong their lives with their loved ones. However, early PD diagnosis and the required early treatments are made possible by data-driven identification models. Researchers can improve patient quality of life and speed up therapy by optimizing diagnostic accuracy with data analytics and technology. In this paper, proposed work different machine learning algorithms like Support Vector Machine, Decision Tree, Random Forest, K Nearest Neighbor algorithms are applied, in that SVM and Random Forest show better performance for three models, yielding accuracy levels of 97%, 86.67%, and 76.65% for the Voice, Spiral, and Wave models, respectively.

Keywords: Machine learning algorithms, Random forest, Spiral and wave drawing, Support vector machine, Voice dataset.

DOI: 10.53894/ijirss.v8i3.7112
Funding: This study received no specific financial support.
History: Received: 19 March 2025 / **Revised:** 22 April 2025 / **Accepted:** 24 April 2025 / **Published:** 16 May 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Millions of people worldwide are impacted by Parkinson's disease (PD), a serious global health concern. Although MRIs, PET scans, and mini-mental state tests are commonly used diagnostic methods, they still have limitations that make it difficult to diagnose Parkinson's disease (PD) in its early stages. These methods are frequently intrusive, time-consuming, and may not accurately identify symptoms in their early stages, which can delay diagnosis and reduce therapy efficacy in later stages. Creating a unique prediction model for PD detection is the goal of this study in order to address these issues.

The model will incorporate information from voice recordings, spiral and wave drawings, and other sources using machine learning approaches to produce a diagnostic tool that is more effective and user-friendly. The major goal is to increase diagnostic precision so that health professionals can identify Parkinson's disease (PD) earlier and with greater accuracy. Among disorders of the nervous system, PD poses a serious challenge due to its progressive nature, which can occasionally lead to a delayed diagnosis and decreased treatment efficacy in later stages.

Our goal is to revolutionize the way Parkinson's disease is managed by utilizing early identification and management. Developing a prediction model, investigating biomarkers, exploring novel therapeutic approaches, and optimizing early therapy are some of our objectives as we sort through the nuances of this complex illness. This massive effort aims to increase scientific understanding, develop a proactive paradigm for treating Parkinson's disease, examine the financial effects, and promote clinical trial participation.

The PD presents a number of difficulties since treatment efficacy is hampered by a late diagnosis. Improving outcomes and revolutionizing PD care depend heavily on early identification. To address this, a multidisciplinary strategy is needed to handle complexity and overcome diagnostic delays. To combat the slow onset of the PD and facilitate timely interventions that can transform treatment success, early detection programs are essential. Utilizing a variety of datasets and cutting-edge technology presents hitherto unseen possibilities for early Parkinson's disease identification, supporting the creation of innovative diagnostic instruments. Timely identification enables customized therapies that improve patient outcomes and quality of life. Prioritizing early intervention measures has a wider societal impact, since proactive early detection initiatives also seek to lessen the financial and social burdens associated with severe PD.

This research is motivated by the urgent need to solve the intricate issues that Parkinson's disease (PD) presents, as well as the revolutionary potential of early identification and intervention. Given the disease's delayed onset and the ensuing delays in diagnosis, our comprehensive approach encompasses a range of objectives. Our objectives are to generate a sophisticated predictive model using cutting-edge machine learning techniques, investigate potential biomarkers, and increase the model's sensitivity for early-stage Parkinson's disease detection.

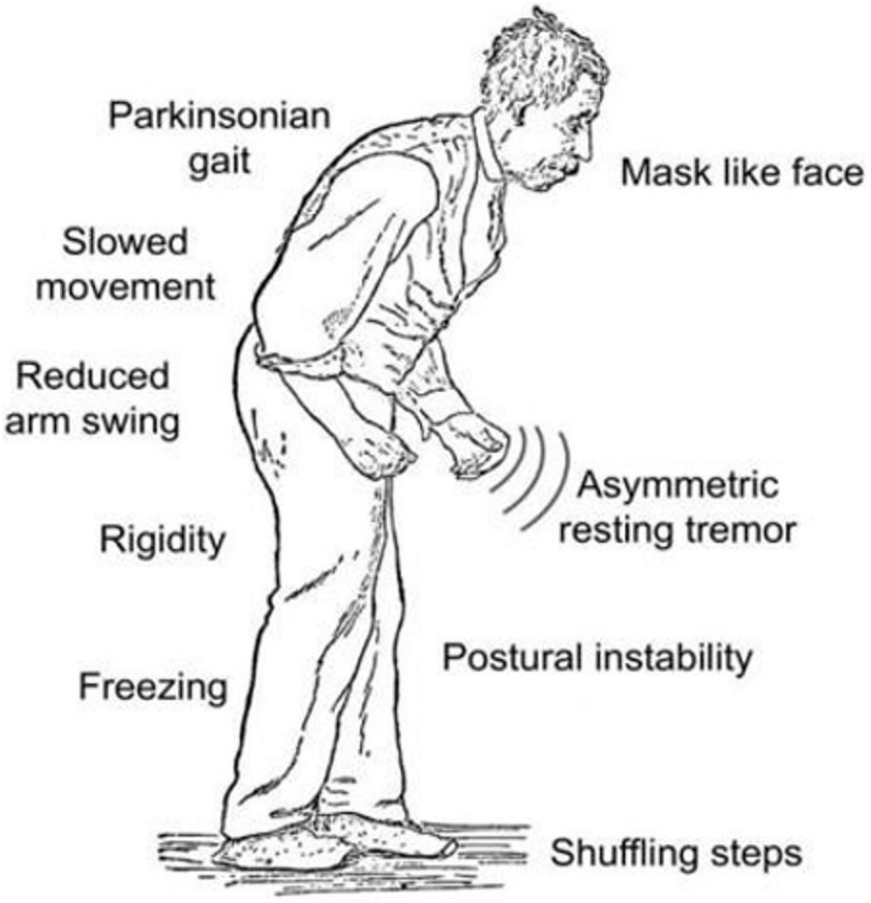

Figure 1.
Parkinson's Disease Symptoms.

Additionally, we strive to investigate cutting-edge therapeutic methods and strengthen timely, targeted actions. In addition, we assess the financial implications, promote clinical trial participation, advance scientific understanding, and create a proactive paradigm shift in Parkinson's disease therapy. Our diverse objectives encompass not just improving patient outcomes but also advancing scientific understanding of the intricacies of Parkinson's disease, medical research, and social well-being.

2. Literature Survey

The survey aims to conduct a comprehensive synthesis of scholarly literature, encompassing research on biomarker identification, improvements in predictive modeling, and early detection methods for Parkinson's disease (PD).

The current prediction method uses the following preparation techniques: input image scaling, contrast enhancement for spiral drawings, histogram equalization, pairwise correlation, and exploratory data analysis for the voice dataset.

This research employs the following methods: VGG16-CNN, k-Nearest Neighbors classifier, Logistic Regression classifier for the Spiral dataset, and Pattern Recognition Fuzzy C-Means (FCM) clustering for the voice dataset. The Parkinson's Telemonitoring dataset, PPMI, and the Parkinson's Disease Voice and Speech dataset were among the datasets taken from the UCI Machine Learning libraries.

Roobini, et al. [1] present a novel deep learning method for voice and speech identification that uses dopaminergic imaging indicators, cerebrospinal fluid, olfactory loss, and rapid eye movement to diagnose Parkinson's disease early on. The higher detection performance of the suggested model is demonstrated by a comparison between the proposed deep learning model and 12 machine learning and ensemble learning approaches based on relatively modest data, including 401 early patients and 183 individuals. which, in addition to having an average accuracy of 96.45%, featured a boosting approach for the purpose of detecting Parkinson's disease. In order to enhance the quality of the data, Alniemi and Mahmood [2] developed a model that included several data pre-processing techniques, including data standardization, multicollinearity diagnosis, and dimensionality reduction approaches. For the purpose of classifying Parkinson's disease, several machine learning (ML) classifiers were employed, including k-nearest neighbor, support vector machines, random forest, AdaBoost, and logistic regression.

Cross-fold validation, grid search, and hyperparameter tuning were utilized in this investigation to optimize the classifiers' performance and maintain the unbalanced dataset's class distribution. SVM achieved the highest accuracy of 94.10% among all the ML methods. Nagarathna and Kusuma [3] suggested a deep learning method for voice recordings. They used a ResNet architecture that was initially intended for image classification and addressed vowels with sustained phonation for detection.

Kaggle [4] presented a comprehensive system architecture for the identification of spiral and wave drawings that can be used to analyze spiral and wave drawing patterns in both healthy individuals and patients with Parkinson's disease. They employed two different convolutional neural networks (CNNs) in their research, both of which were dedicated to the study of drawing patterns, namely spiral and wave sketches. Additionally, a meta-classifier's prediction probabilities were trained using ensemble voting techniques, which enhanced the classifier's prediction strength.

With the help of this technique, it was simpler to create a weighted prediction based on spiral and wave sketch assessments. The model was trained on a dataset with 55 patients and, after extensive training, attained an impressive accuracy of 93.3%. Based on the drawing patterns, the generated system was able to differentiate between healthy individuals and those with Parkinson's disease, thanks to the output model. Using the Federal University of Uberlândia's NIATS dataset, which included 500 photos of spiral drawings created by both healthy individuals and Parkinson's patients, [5] a CNN model was trained on 500 spiral drawings from patients with Parkinson's disease.

1. 1. Using voice recordings and a ResNet architecture, Marek Wodnizki, Andrzej Skalski, and colleagues [3] *developed a deep learning model.*

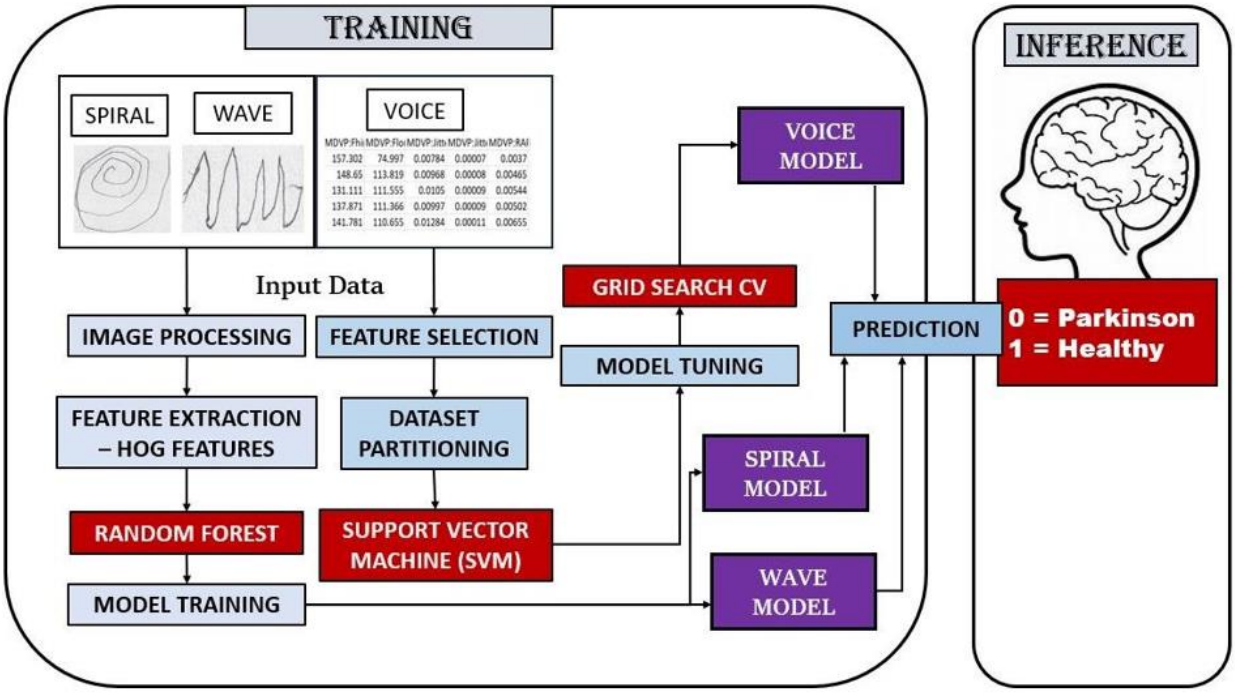

Figure 2.
The proposed data flow diagram.

2. Learning method for identifying Parkinson's disease in images. With a comparable accuracy of 90%, they used data from 100 patients, each of whom was recorded three times.

Gaps identified [6] in literature about PD are listed below.

- Using ensemble learning with additional classification techniques, such as kernel dictionary learning, SVM and improved k-NN; the models were trained on a limited Vocal test and Spiral test data.
- To make the models more robust and reliable, additional data from the patients suffering from PD during the early stages of the disease from different parts of the world are needed
  Expand the dataset to include more cases within each class of severity and combine
  Patient voice data with further patient attributes such as handwriting and gait features.
  Making the system functional in reliable production environments and real-world situations.
- Parkinson's disease diagnosis is often based on the patient's history, assessments, and observations from the clinical examinations and questionnaires about particular symptoms because currently there are no approved, definitive biomarkers or specific neuroimaging findings. The use of ensemble learning with supplementary classification methods, such as enhanced k-nearest neighbors and SVMs.

3. Methodology

The methodology used in this work incorporates modules for data analysis, feature extraction, and patient classification according to health status—that is, whether the patient is well or has been diagnosed with Parkinson's disease.

3.1. Data Acquisition

Matrix column entries (attributes):

- **name** - ASCII subject name and recording number
- **MDVP:Fo(Hz)** - Average vocal fundamental frequency
- **MDVP:Fhi(Hz)** - Maximum vocal fundamental frequency
- **MDVP:Flo(Hz)** - Minimum vocal fundamental frequency
- **MDVP**:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP - Several measures of variation in fundamental frequency
- **MDVP**:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Several measures of variation in amplitude
- **NHR, HNR** - Two measures of the ratio of noise to tonal components in the voice
- **status** - The health status of the subject (one) - Parkinson's, (zero) - healthy
- **RPDE, D2** - Two nonlinear dynamical complexity measures
- **DFA** - Signal fractal scaling exponent
- **spread1, spread2, PPE** - Three nonlinear measures of fundamental frequency variation

Figure 3.
Voice data attributes.

3.1.1. Voice Dataset

The voice dataset is received from the UCI Machine Learning Repository [7]. It is composed of 31 people, 23 of whom have been diagnosed with Parkinson's disease, and a wide variety of biological speech measurements. The 195 voice recordings that these individuals made are represented by the "name" column in the table, where each row corresponds to a particular voice measure. The main way to distinguish between people without Parkinson's disease (PD) and those who have it is to look at the "status" column, which has values of 0 for healthy and 1 for PD. The data is in ASCII CSV format. A single voice recording instance is represented by each row in the CSV file. About six recordings are made for each patient; the patient's name is shown in the first column. The details about the features of the voice dataset used in our work are listed in Figures 3 and 4 respectively.

Attribute	Description	Target 0		Target 1	
		Maximum	Minimum	Maximum	Minimum
MDVP:Fo(Hz)	Average vocal fundamental frequency	260.105	110.739	223.361	88.333
MDVP:Fhi(Hz)	Maximum vocal fundamental frequency	592.03	113.597	588.518	102.145
MDVP:Flo(Hz)	Minimum vocal fundamental frequency	239.17	74.287	199.02	65.476
MDVP:Jitter(%)	Several measures of variation in fundamental frequency	0.0136	0.00178	0.03316	0.00168
MDVP:Jitter(Abs)		0.00008	0.000007	0.00026	0.00001
MDVP:RAP		0.00624	0.00092	0.02144	0.00068
MDVP:PPQ		0.00564	0.00106	0.01958	0.00092
Jitter:DDP		0.01873	0.00276	0.06433	0.00204
MDVP:Shimmer	Several measures of variation in amplitude	0.04087	0.00954	0.11908	0.01022
MDVP:Shimmer(dB)		0.405	0.085	1.302	0.09
Shimmer:APQ3		0.02336	0.00468	0.05647	0.00455
Shimmer:APQ5		0.02498	0.00606	0.0794	0.0057
MDVP:APQ		0.02745	0.00719	0.13778	0.00811
Shimmer:DDA		0.07008	0.01403	0.16942	0.01364
NHR	Two measures of the ratio of noise to tonal components in the voice	0.10715	0.00065	0.31482	0.00231
HNR		33.047	17.883	29.928	8.441
RPDE	Two nonlinear dynamical complexity measures	0.663842	0.25657	0.685151	0.263654
D2		0.785714	0.62671	0.825288	0.574282
spread1	Three nonlinear measures of fundamental frequency variation	-5.198864	-7.964984	-2.434031	-7.120925
spread2		0.291954	0.006274	0.450493	0.063412
PPE		2.88245	1.423287	3.671155	1.765957
DFA	Signal fractal scaling exponent	0.252404	0.044539	0.527367	0.093193

Figure 4.
Minimum and Maximum values for Voice data attributes.

3.1.2. Spiral Dataset: The Spiral and wave dataset of Parkinson's Disease are Collected from Kaggle [4].

It consists of drawings of spirals created by a group of people: healthy controls and those with Parkinson's disease diagnoses. There are 50 images in the testing set and 100 images in the training set for this dataset. In order to identify variations in motor control and support the creation of diagnostic tools for ML models of Parkinson's disease, these images are evaluated.

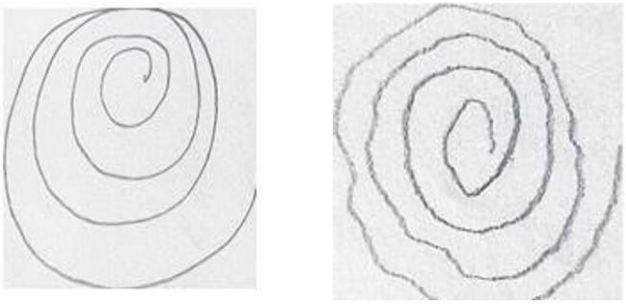

Figure 5.
Healthy spiral drawing and Parkinson's disease affected person spiral drawing.

3.1.3. Wave Dataset

Drawings of individuals with Parkinson's disease and healthy controls make up the composition. Like the spiral dataset, this training set also contains 100 photos.

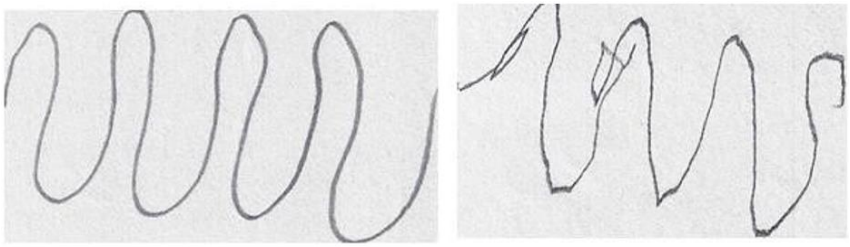

Figure 6.
Healthy spiral drawing and Parkinson's disease affected person wave drawing

And the testing set's fifty photos. They support the evaluation of motor control, the identification of traits associated with the condition, and the improvement of the validity of results pertaining to the diagnosis of Parkinson's disease.

3.2. Data Preprocessing

1) Voice Data Preprocessing: The voice feature values are preprocessed by either manually entering them into the system or automatically entering them. The procedure involves selecting the features that are most pertinent, most influential, and have a major impact on the model's accuracy. It is possible to successfully identify important vocal elements or characteristics, such as pitch variation or loudness, that aid in the accurate diagnosis and therapeutic planning process by using machine learning preprocessing techniques and patterns suggesting the presence of illness.

2) Image preprocessing: Images must go through the required pretreatment steps in order to undergo the necessary changes to improve and standardize their quality before feature extraction can begin. In order to retain the structural components, the images are first converted from color to grayscale. To maintain analytical consistency, the dataset is then downsized to a uniform size of 200x200 pixels across the dataset. Lastly, the images are divided into portions for background noise and foreground based on the intensity levels of the pixels by adjusting the pixel intensity with a single threshold. In a 2D image, represent a pixel as f(x,y) where the horizontal and vertical coordinates are denoted by x and y, respectively. A global threshold is a distinct threshold T that is applied to each pixel in the image. Selecting the ideal threshold t to divide the foreground—that is, the object and backdrop—is a simple method of detecting or extracting items from the background, and it is explained in detail below.

3.2.1. Thresholding

Otsu's method is used to convert grayscale images into binary format, which facilitates feature extraction. Otsu's method aims to find a threshold T that minimizes the within-class variance of pixel intensities for a binary image. It is computed from the histogram of pixel intensities. The basic idea is to evaluate the optimal threshold that divides the histogram into two classes, $C1$ (background) and $C2$ (foreground), such that the sum of within-class variances is minimized.

i. Histogram and Probability Distribution: Given a grayscale image, create a histogram of pixel intensities. Let $P(i)$ represent the probability (or relative frequency) of pixel intensity i.

ii. Class Probabilities: Compute the cumulative probabilities for the two classes at a threshold T

$w_1 = \sum_{i=0}^{T} p_i$
$w_2 = \sum_{i=T+1}^{255} i * p_i$

iii. Class Means: Compute the class means for the two classes at threshold T:

iv. Between-Class Variance: The between-class variance is given by:

$\sigma^2 = \omega_1 . \omega_2 . (\mu_2 - \mu_1)2$

v. i. Optimal Threshold: Find the threshold T that maximizes the between-class variance:

$T^* = argmax_T\ \sigma^2$

Preprocessing methods are ultimately essential for increasing computational efficiency since they improve image quality and prepare the dataset for precise feature extraction.

3.3. C Feature Extraction of images

The Histogram of Oriented Gradients (HOG) feature extraction approach is essential for diagnosing Parkinson's disease because it extracts important texture and shape information from pre-processed images. The HOG features are useful for extraction because they are effective at drawing attention to important patterns in photographs. Following preprocessing, the recovered HOG features are utilized to train a Random Forest Algorithm machine learning model.

i. Voice Features:

- Extract key voice features such as jitter, shimmer, HNR (Harmonic-to-Noise Ratio), and others.
- Select the most informative features based on correlation analysis and domain knowledge.

ii. Image Features

- Apply Histogram of Oriented Gradients (HOG) to extract edge-based features from spiral and wave images.
- The gradient for a pixel in an image can be calculated using the partial derivatives of the image intensity in both horizontal and vertical directions. Given an image I (x, y), the gradient components **Gx** and **Gy** can be computed as:

  $G_x = I\ (x + 1,\ y) - I\ (x - 1,\ y)$

  $G_y = I\ (x,\ y + 1) - I\ (x,\ y - 1)$
- The magnitude of the gradient G (x, y) is given by:

$$|G\ (x,\ y)| = \sqrt{G_x^2 + G_y^2}^2$$

- The orientation of the gradient \theta (x, y) is given by:

$$\theta\ (x,\ y) = \arctan\left(\frac{G_x}{G_y}\right)$$

- Choose a set of features for training, aiming for a balance between dimensionality and predictive power.

Training gives the model the ability to distinguish between people with Parkinson's disease and healthy patients. In order to identify new data instances, decision boundaries are created by analyzing the patterns and characteristics present in HOG features. Using the recovered features, the model may then be trained to accurately identify Parkinson's disease based on the visual characteristics that the HOG feature descriptors capture. This method enables the precise and automatic identification of Parkinson's disease from the input data, which aids in the disorder's early detection and effective treatment.

3.4. D Model Training

- Data Splitting: Split the dataset into training and testing sets (e.g., 80/20 ratio) to evaluate model performance.
- Model Selection: Use Support Vector Machines (SVM) with a linear kernel for initial training.
- Optimize model parameters with Grid Search CV to find the best configuration (e.g., C, gamma, kernel).
- Consider using ensemble methods like Random Forest for additional robustness.
- Model Training and Evaluation: Train the SVM model on the training dataset.
- Evaluate the model on the test dataset using metrics like accuracy, F1-score, precision, and recall.
- Generate a classification report and ROC curves for further analysis.

3.5. E Model Deployment

- Flask Framework Integration: Develop a front-end application using Flask to provide a user interface for doctors and patients.
- Set up routes to handle image and voice data uploads and return model predictions.
- Implement secure file handling and result display (e.g., showing processed images with prediction results).
- Testing and Validation: Conduct user testing to validate the front-end interface.
- Ensure predictions are displayed accurately and results are easy to understand.

3.6. F Model Serialization

- Model Saving: Serialize the trained models using a library like joblib or pickle for persistent storage.
- Store serialized models securely for later use or deployment in other environments.
- Continuous Monitoring and Updates
- Implement mechanisms to monitor model performance in production.

- Plan for periodic retraining and updates to improve accuracy and adapt to new data patterns.

This algorithm captures the end-to-end process for building and deploying a Parkinson's disease detection system using multiple data sources and machine learning techniques.

3.7. G Data Structuring for Spiral and Wave model

In order to organize extracted features from photos into a format that can be used to train a model, the feature matrix is crucial. This matrix serves as the foundation for the machine learning algorithm, which searches the data for patterns and relationships. Every data point contains a label that states whether the individual is well or has received a Parkinson's disease diagnosis. Next, the algorithm uses the input data to forecast with the aid of these labels, which offer essential information for supervised learning. This type of architecture aids in the model's ability to discriminate between multiple classes, which helps with precise disease detection. The organized method expedites this procedure, enhancing the model's ability to generalize the fresh data and increasing the precision of the diagnosis of Parkinson's disease.

3.8. H Machine Learning Models

1) Voice models: Different machine learning methods are used in voice data categorization to distinguish between individuals who are healthy and those who have Parkinson's disease.

Table 1.
Comparison table for different Machine Learning models for Voice data classification.

Metric	DT	RF	LR	SVM	NB
Accuracy	0.8474	0.9491	0.8305	0.9661	0.7627
F1-Score	0.8421	0.9411	0.7826	0.9600	0.6500
Recall	0.9230	0.9230	0.6923	0.9230	0.5000
Precision	0.7741	0.9600	0.9000	1.0000	0.9285
R2-Score	0.3811	0.7937	0.3123	0.8624	0.0372

(DT-Decision Tree, RF-Random Forest, LR-Logistic Regression, SVM-Support Vector Machine, NB- Naive Bayes)

The Decision Tree model exhibits some effectiveness in categorization, with a moderate accuracy of 84.75%. The Random Forest model stands out as a promising option for ensemble learning in disease classification, with its impressive accuracy of 94.92%. At 83.05% accuracy, Logistic Regression performs admirably, but it is not as good as more intricate models. With an accuracy of 96.61%, the Support Vector Machine (SVM) is the top performer and demonstrates its resilience in the classification of Parkinson's disease.

However, Naïve Bayes performs worse, with an accuracy of 76.27%, indicating that it may not be able to capture complex patterns of sickness. SVM appears to be the most advantageous choice in general due to its increased classification accuracy for Parkinson's disease.

GridSearchCV is used by the voice data model to fine-tune its hyperparameters. It carefully examines a preset grid of parameter values using cross-validation, optimizing model parameters to boost efficiency and generalizability. By selecting the set of parameters that yield the best model performance, this thorough search helps to improve the accuracy and reliability of the machine learning models that are used to categorize people with Parkinson's disease from those in good health using voice data.

2) Spiral and Wave models: The Random Forest and XGBoost algorithms' accuracy ratings in identifying the
The table that follows displays the wave and spiral models. In both tasks, Random Forest performs better.

Table 2.
Accuracy table for Spiral and Wave classification.

Algorithms	Spiral model	Wave model
Random Forest	86.86%	76.65%
XGBoost	73.33%	73.33%

Random Forest Classifier is preferred over XGBoost and Neural Networks mostly due to its higher accuracy in illness classification tasks. Random Forest makes optimal use of a set of decision trees to lessen overfitting and increase prediction accuracy. It reduces the possibility of lowering the bias and variance associated with individual trees by aggregating the predictions from multiple trees, which results in predictions that are more reliable and consistent. Random Forest is a suitable option for this application in identifying diseases since it is interpretable, robust against overfitting, very accurate, and easy to use.

3.9. I Prediction of the Models

The model is evaluated on a different set of images and speech characteristics after training to assess how well it can recognize Parkinson's disease. The model's capacity to discriminate between individuals with Parkinson's disease and healthy ones is now being assessed. In this stage, the test results are visualized. For example, forecasts can be shown in an easy-to-understand style, or results can be exported for medical review. By examining the outcomes, researchers and medical professionals can assess the model's viability and efficacy for actual use in Parkinson's disease detection.

Depending on whether Parkinson's disease symptoms are present or absent, the final result classifies the input data as

either "0–Healthy" or "1–Positive."

4. Results

Whether or not a person has Parkinson's disease was successfully determined using the suggested method, based on the testing and identification of the model using the integrated method for voice, wave, and spiral data. To achieve the aforementioned outcomes, the suggested work utilized an Intel ® Core TM i7-8750H, 16GB of RAM, a 64-bit operating system, and an NVIDIA GPU.

Determining whether an individual has Parkinson's disease or not was successfully accomplished using the suggested method.

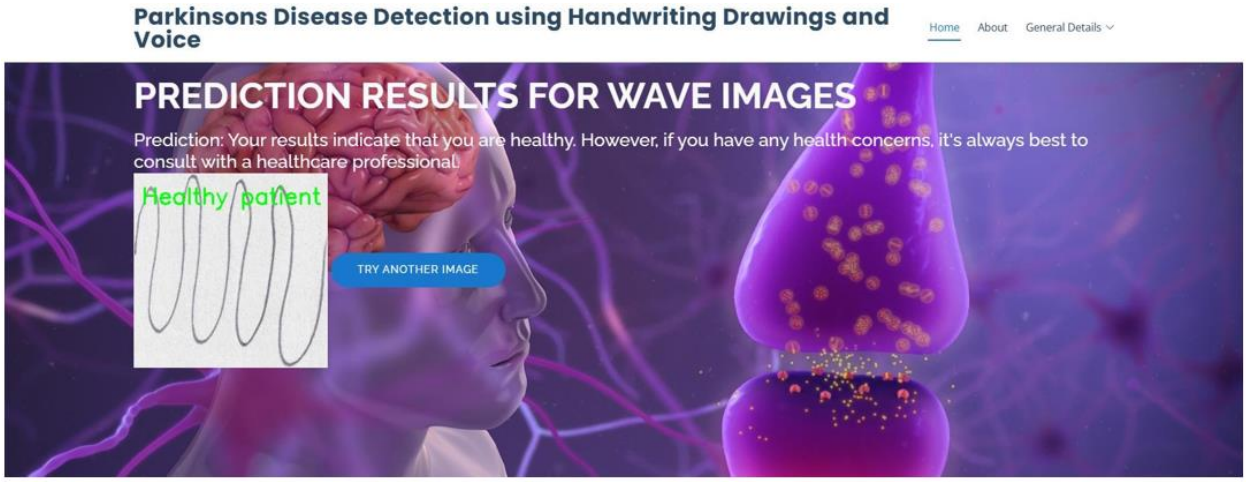

Figure 7.
Results from Wave drawing.

Figure 8.
Results from Spiral drawing.

Use a combined approach to Spiral, Wave, and Voice data for testing and detection of the model.

Figure 9.
Results from Voice data.

The spiral model produced an accuracy of 86.67% when used to forecast the spiral drawing of a Parkinson's patient, as seen in Figure 8.

Figure 9 illustrates how speech data was predicted using the speech model with a 97% accuracy rate based on the attribute values entered.

The AUC-ROC curve for the SVM algorithm used in the Voice model to identify Parkinson's disease is displayed in Fig. 10.

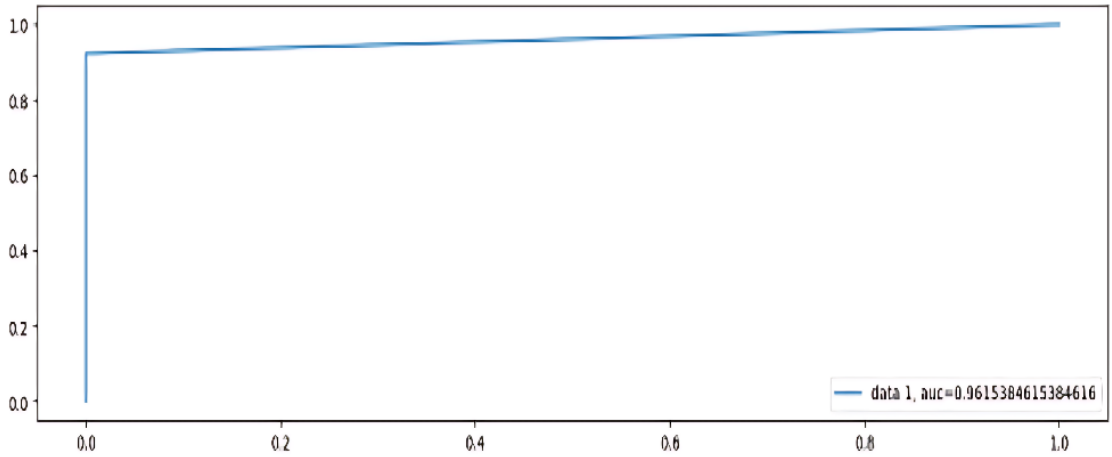

Figure 10.
AUC- ROC curve for Voice model.

5. Conclusion

The goal of the entire project is to advance our knowledge of Parkinson's disease and our capacity to treat it by using a thorough approach to early identification and intervention. The amalgamation of an extensive array of information, sophisticated machine learning models, and investigations into biomarkers and therapeutic strategies has facilitated novel comprehensions of the intricate characteristics of Parkinson's disease.

Subsequent improvements and developments will focus on enhancing the machine learning model by including other data sources, utilizing wearable technology, and fostering collaborations with medical specialists. Future advancements can be facilitated by the combination of ongoing research into novel biomarkers and therapeutics with longitudinal investigations.

6. Future Enhancements

Prospective developments in the identification of Parkinson's disease (PD) are intended to use state-of-the-art technologies and techniques to improve patient outcomes and diagnostic precision. A thorough picture of the course of Parkinson's disease (PD) can be attained by combining several data sources, such as voice and handwriting data, with neuroimaging scans and wearable sensors. Additionally, investigating cutting-edge machine learning methods, such as deep learning algorithms, has the potential to uncover minute patterns that indicate the beginning and course of Parkinson's disease (PD) and enhance diagnostic accuracy.

Customizing treatment approaches to individual needs and following the evolution of diseases over time are two ways that longitudinal data analysis offers prospects for personalized medicine. Wearable technology and smartphone apps are examples of real-time monitoring solutions that may help with prompt intervention and proactive symptom management. To guarantee that these advances are seamlessly incorporated into clinical practice, novel diagnostic tools must be developed in collaboration with healthcare professionals. Only then can the quality of life of patients with Parkinson's disease (PD) be improved through novel diagnostic tools that meet stringent requirements.

References

[1] M. Roobini, Y. R. K. Reddy, U. S. G. Royal, A. K. Singh, and K. Babu, "Parkinson's disease detection using machine learning," in *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, 2022: IEEE, pp. 1-6.

[2] O. Alniemi and H. F. Mahmood, "Convolutional neural network for the detection of Parkinson disease based on hand-draw spiral images," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 1, pp. 267-275, 2023. https://doi.org/10.11591/ijeecs.v30.i1.pp267-275

[3] C. Nagarathna and M. Kusuma, "Comparative study of detection and classification of Alzheimer's disease using hybrid model and CNN," in *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, 2021, vol. 1: IEEE, pp. 43-46.

[4] Kaggle, "Parkinson disease spiral drawings using digitized graphics tablet. Kaggle," Retrieved: https://archive.ics.uci.edu/ml/datasets/Parkinson+Disease+Spiral+Drawings+Using+Digitized+Graphics+Tablet, 2020.

[5] L. Akter, "Early identification of Parkinson's disease from hand-drawn images using histogram of oriented gradients and machine learning techniques," in *2020 Emerging Technology in Computing, Communication and Electronics (ETCCE)*, 2020: IEEE, pp. 1-6.

[6] G. Himaja, C. R. Nagarathna, A. Jayasri, and K. M. Kundan, "Prediction of Parkinson's disease using handwriting analysis and voice dataset- a review," *Journal of Innovative Image Processing*, vol. 6, no. 2, pp. 118-132, 2024.

[7] UCI Machine Learning Repository, "Parkinson's disease dataset. UCI Machine Learning Repository," Retrieved: https://archive.ics.uci.edu/dataset/174/parkinsons, 2020.