



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

A Brief Review on Parkinson's Diseases detection with various Machine Learning Techniques

¹Ghanshyam G. Parkhede, ²Dhananjay S Deshpande

¹Saraswati College, Shegaon, Maharashtra, India. E-mail: ghanshyamparkhede@gmail.com

²MBAESG, School of Management, Ajeenkya D Y Patil University, Pune, Maharashtra, India, E-mail: drdeshpande.dhananjay@gmail.com

Abstract. Parkinson's disease (PD) is a chronic neurodegenerative illness that affects the brain and central nervous system, leading to issues with pain, mobility, mood, and sleep. Early and precise diagnosis is vital for effective medical intervention and enhancing the quality level of life for patients. This review provides an extensive overview of current PD detection methods, emphasizing the integration of neuroimaging techniques, clinical data (bio markers) and advanced computational algorithms. A wide range of neuroimaging modalities are examined by highlighting their roles in identifying structural and functional brain changes linked to PD. The review also explores the potential of datasets such as facial expressions analysis, Bio Chemical Markers, voice recordings, etc. The review discusses various stages of data processing, including preprocessing, segmentation, and feature extraction, essential for improving the accuracy and efficiency of diagnostic models. The application of machine learning, deep learning, and transfer learning models for PD classification and prediction is reviewed, focusing on feature selection, model optimization, and the utilization of large, diverse datasets. AI systems can provide reliable and accurate PD identification, enabling more efficient and prompt therapeutic treatments, by utilizing these sophisticated algorithms and a variety of datasets.

Keywords: Parkinson's disease, dataset, machine learning, transfer learning, segmentation.

1 Introduction

This Parkinson's disease (PD) is a slowly progressing neurodegenerative disorder marked that affects motor function by a gradual decrease of dopamine-producing neurons in the brain, mostly in the substantia nigra [1,2]. Many motor symptoms result from this decrease in dopamine, a crucial neurotransmitter involved in controlling movement. These include tremors, muscle rigidity, bradykinesia (slowness of movement), and difficulties with balance and coordination. Over time, these symptoms become more pronounced, making everyday tasks like walking, dressing, and eating increasingly challenging, which can significantly reduce a person's independence. Apart from its impact on motor function, PD includes a broad spectrum of nonmotor symptoms, including mood disorders, sleep problems, cognitive loss, and autonomic dysfunction [3-5].

The non-motor components of Parkinson's disease contribute significantly to its complexity and can have a significant impact on a person's overall quality of life. The diagnosis of Parkinson's disease primarily relies on a clinical evaluation performed by neurologists. These professionals assess symptoms using established rating tools, such as the Movement Disorder Society - Unified Parkinson's Disease Rating Scale and the Hoehn and Yahr scale [6-9]. Through meticulous analysis, neurologists gauge the severity and progression of motor symptoms, essential for accurate diagnosis and treatment planning. These rating scales serve as essential tools in the clinical toolkit, facilitating consistent and objective evaluation across patients. Their utilization aids in identifying characteristic features of Parkinson's disease, aiding in distinguishing it from other movement disorders or neurological conditions[10].

Diagnosing PD can be challenging, particularly early on, due to the variability and overlap of symptoms with other movement disorders. Additionally, there are currently no definitive biomarkers for PD, necessitating reliance on clinical judgment and observation. As a result, misdiagnosis can occur, leading to suboptimal management and treatment outcomes [11,12]. As a result, more approachable but as of yet unproven techniques for early Parkinson's disease identification are required.

Early detection of PD is vital for timely treatment and management to alleviate symptoms and potentially slow down disease progression. Advancements in medical imaging technologies, such as EEG, MRI, and PET, alongside sophisticated analyses of motor symptoms like gait, handwriting, and speech, are transforming our ability to detect and diagnose PD at its earliest stages. EEG offers insights into neural oscillations and cognitive dysfunction, while MRI provides detailed structural and

functional information about PD-related brain changes. PET imaging aids in detecting metabolic and dopaminergic dysfunction, crucial for accurate diagnosis and treatment monitoring [13-15]. Additionally, quantitative analyses of gait, handwriting, and speech offer objective measures of motor symptoms, facilitating early detection and personalized interventions for PD patients. These multidimensional approaches not only enhance our understanding of PD pathophysiology but also hold promise for optimizing therapeutic strategies and improving patient outcomes [16, 17].

The integration of neuroimaging modalities with advanced analyses of motor symptoms signifies a important phase in the early prediction of Parkinson's Disease. By combining EEG, MRI, and PET with quantitative assessments of gait, handwriting, and speech, clinicians gain comprehensive insights into PD pathology and disease progression. This comprehensive approach allows for the development of personalized treatment plans that cater to the specific needs of each patient, resulting in better symptom management and an improved value of life for persons with Parkinson's Disease. Moreover, ongoing research efforts in this interdisciplinary field promise to further refine diagnostic accuracy, monitor disease progression, and develop novel therapeutic targets, ultimately advancing our ability to combat Parkinson's Disease effectively [18].

Conventional diagnostic methods depend on clinical evaluations, which may lack the sensitivity needed to identify changes during the early time of the disease. However, new advancements in machine learning and deep learning have created promising opportunities to enhance the detection and diagnosis of Parkinson's Disease. These algorithms have shown exceptional ability in interpreting intricate medical data, particularly neuroimaging scans, genetic profiles, and clinical data, to extract meaningful patterns and assist in disease diagnosis [19, 20]. By learning from various features extracted from neuroimaging data, such as structural abnormalities, metabolic changes, and disruptions in functional connectivity, ML and DL algorithms can identify important biomarkers that indicate PD pathology. Deep Learning holds tremendous potential for refining the detection of Parkinson's Disease (PD) by extracting meaningful features from neuroimaging data and clinical assessments.

Deep learning algorithms such as convolutional networks, recurrent networks, and advanced types like Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM), offer unprecedented capabilities in analyzing complex neuroimaging data and clinical assessments. By meticulously analyzing pixel-level features and hierarchical representations within MRI images, CNNs can determine abnormalities in brain morphology characteristic of PD, providing clinicians with valuable insights for early diagnosis and treatment planning. Recurrent Neural Networks includes advanced architectures offer unparalleled capabilities in processing sequential data from wearable sensors for assessing PD-related motor symptoms like gait abnormalities. RNNs can effectively capture temporal dependencies in sensor data, allowing for the detection of subtle variations in gait patterns indicative of PD progression. With LSTM and Bi-LSTM, RNNs can model long-term dependencies and bidirectional information flow, enhancing their capability to detect significant features from input data and providing clinicians with valuable insights into the progression of PD-associated motor impairments.

2 Discussion on datasets and existing work on PD

Current work includes:

- A comprehensive analysis of neuroimaging techniques (MRI, PET, EEG) and their role in identifying structural and functional brain changes in Parkinson's Disease (PD).
- Exploration of diverse, non-invasive datasets such as handwriting samples, voice recordings, gait analysis, EEG, and sensor data for enhancing PD detection.
- Emphasis on the application of machine learning, deep learning, and transfer learning for PD classification, focusing on feature selection, model optimization, and the use of large datasets to improve diagnostic accuracy.

The process of diagnosing Parkinson's Disease (PD) using MRI images involves several key steps. Initially, MRI images, including T1 and T2-weighted sequences, are acquired from databases or clinical sources. These images undergo preprocessing, which includes segmentation to isolate relevant brain regions, normalization for consistent intensity values, and quality control to ensure they meet analysis standards. Relevant features are then extracted from the segmented images, followed by the development of machine learning models trained on these features to distinguish between PD patients and healthy controls. The models are evaluated based on accuracy and other metrics, and once validated, they can be used to predict PD in new MRI images. Finally, collaboration with healthcare professionals is essential for clinical validation of the findings through trials or studies.

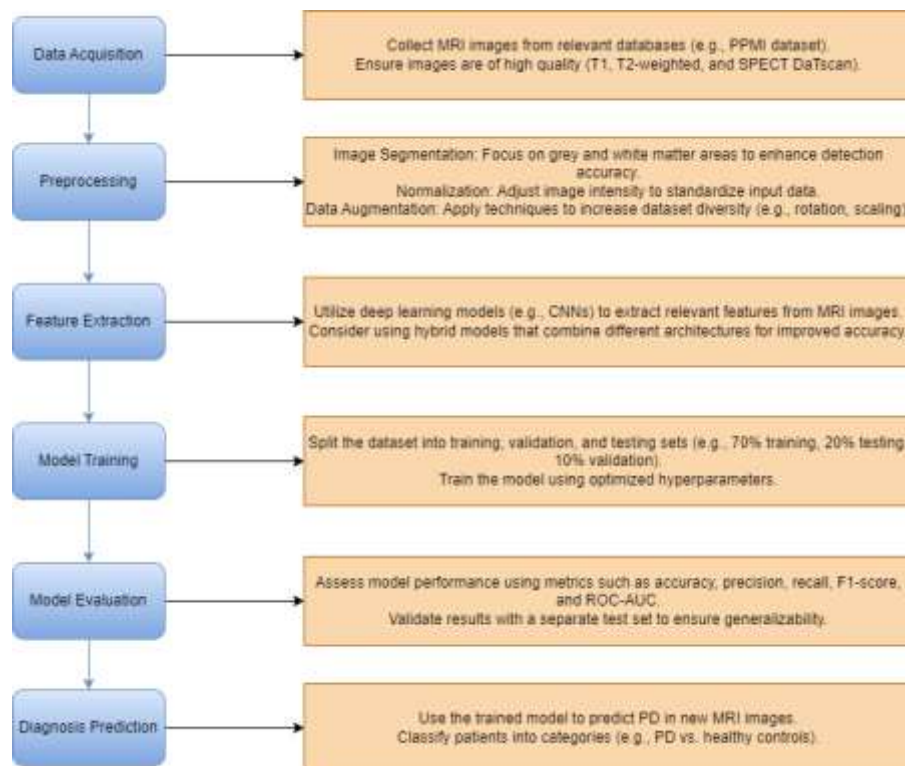


Figure 1 Detecting Parkinson's Disease using MRI Analysis

Figure 1 shows structured approach highlights the systematic analysis of EEG data for detecting Parkinson's disease, emphasizing preprocessing, feature extraction, and classification as critical components of the detection process. Each step is informed by recent studies that utilize advanced techniques in EEG analysis to achieve high classification accuracy for PD detection.

Table 1 MRI Image Analysis

Authors	Work Done
R. Souza, A. Winder, E. A. M. Stanley [1]	The multicentre PD database [1] is used for evaluating Parkinson's Disease. This database includes 1880 T1-weighted MRI scans in which 867 images are PD images and 1013 images are normal images. These samples are collected across 41 different centers. A training set and a test set comprise the dataset, 20% of the dataset is set aside for testing, and the remaining 80% is used for training. During training process, 1478 images are utilized to train the model which consists of 680 samples with PD disease and 798 healthy samples. The training set of 680 disease images holds 418 males, and 262 females, another set of 798 healthy images includes 497 males, and 301 females. The test set contains 402 images which includes 187 PD images and 215 normal samples. There were 124 males and 63 females in the PD images and 138 males and 77 females in the normal healthy images.
M. Camacho, M. Wilms, P [2]	T1-weighted MRI datasets were utilized [2] for Parkinson's Disease prediction. This dataset was created by collecting 2,041 images from 13 studies in which, 1,024 MRI scans were gathered from patients and 1,017 MRI scans were obtained from a healthy people who were matched by age and sex.
S. Shinde, S. Prasad [3]	Shinde et al. [3] done research work with Neuromelanin sensitive magnetic resonance imaging which helps to detect the abnormalities in PD. A collection of 25 healthy samples and 30 PD images are used for training and cross-validation. The 5-fold validation approach is employed to analyze the diseases. The 25 samples are given for testing phase in which 10 samples are healthy and 15 samples are affected by PD
S. Chakraborty, S. Aich [4]	Parkinson's Progression Markers Initiative (PPMI) database is used [4] in which 3T T1-weighted MRI scans are collected from 406 subjects during their baseline visit. Within this dataset, 203 subjects were identified as healthy and the remaining 203 subjects were affected by Parkinson's Disease.
V. Madge, V. S. Fonov [5]	The MRI templates [5] were created from 143 patients. The 126 subjects are PD patients, consisting of 44 females and 82 males, and the patients were aged between 40 and 87 years. Additionally, the dataset includes 17 healthy samples that contains 13 females and 4 males, and the people were aged between 39 and 84 years.

T. M. Tassew, N. Xuan, and B. Chai [6]	For an MRI study [6], T2-weighted images in T1 anatomical space were obtained from the PPMI dataset. The research work was done by utilizing the data from 450 people with early-stage PD and 103 healthy control subjects. All the people were aged between 50 and 79 years. To avoid class imbalance problem, 347 MRI images of healthy control subjects were collected from the IXI dataset which are publicly available.
M. Camacho, M. Wilms [7]	A multimodal MRI dataset was utilized [7] which incorporates both T1-weighted and DTI scans from patients and matched healthy controls. Totally 1300 subjects were collected from eight different studies and it includes 632 patients and 668 healthy control subjects, Ariz et al. utilized the 3-T Trio TIM MR dataset, in which 40 samples are healthy and 39 samples are obtained from PD patients
M. Ariz, R. C. Abad, G. Castellanos [8]	The disease is diagnosed by a movement disorders specialist using the UK Brain Bank criteria. The dataset [8] is obtained from PPMI which includes 500 T2-weighted MRI images for both PD patients and healthy controls.
S. S. Hussain, X. Degang [9]	The PPMI dataset [9] is utilized for PD detection. This dataset contains information of 161 files on various patients, both with and without PD. Magnetic resonance imaging (MRI) data were obtained from a group which includes 29 healthy participants and 69 individuals with PD.
R. Begum, T. P. Kumar [10]	Within the PD group, subjects exhibited varying degrees of disease severity, ranging from mild to moderate. Conversely, all participants within the healthy control group had no prior history of neurological disorders, ensuring a clear distinction between the two study groups. T2-weighted MRI data from 102 PD and 102 normal control subjects were used, [10] sourced from the PPMI dataset. The SV_3Dataset contains 612 slices, including 306 from PD subjects and 360 from healthy controls (HC).

Recent advancements in the identification of Parkinson's disease (PD) have employed electroencephalography (EEG) data analysed through machine learning algorithms. A method employing Support Vector Machine (SVM) models enhances diagnostic accuracy through the analysis of EEG data from Parkinson's Disease patients and healthy controls. This method encompasses extensive feature engineering, hyperparameter optimization, and kernel selection, resulting in improved interpretability and ethical considerations for clinical applications. Research has utilized common spatial pattern methodologies alongside various machine learning classifiers, achieving classification accuracies exceeding 99% in off-medication conditions and approximately 95-98% during medication use. The techniques demonstrate the potential of EEG analysis for the early detection of Parkinson's disease, which may enhance patient outcomes through timely intervention and treatment strategies.

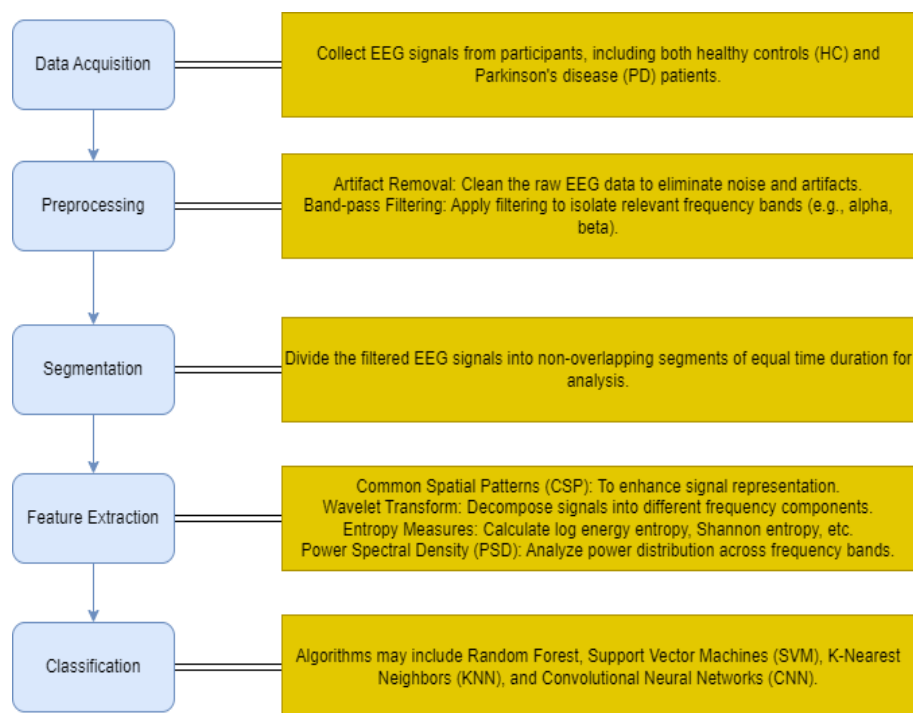


Figure 2 Detecting PD using EEG and Machine Learning

Figure 2 shows structured approach highlights the systematic analysis of EEG data for detecting

Parkinson's disease, emphasizing preprocessing, feature extraction, and classification as critical components of the detection process. Each step is informed by recent studies that utilize advanced techniques in EEG analysis to achieve high classification accuracy for PD detection.

Table 2 EEG Data Analysis

Authors	Work Done
M. Aljalal, S. A. Aldosari [11]	SanDiego dataset and UNM dataset are used with discrete wavelet transform (DWT) methods and entropy measures to spot Parkinson's disease (PD) from EEG signals. The San Diego dataset comprises EEG data of 16 healthy humans and 15 PD patients, with recordings taken both on- and off-medication at a 512 Hz sampling frequency. The UNM dataset contains EEG data from 27 PD patients and 27 healthy individuals, recorded with 64 channels at 500 Hz, both on- and off-medication.
Z. Cao, A. R. John, H.-T. Chen [12]	A self-collected EEG dataset was created by involving seventeen PD affected individuals for the identification of EEG dynamics. EEG data were recorded under various categories by performing timed up-and-go (TUG) tasks, both standard and with voluntary stopping. EEG epochs surrounding FOG and voluntary stopping episodes were analyzed using event-related spectral perturbation (ERSP) to investigate brain dynamics.
C. Chu, Z. Zhang, Z. Song [13]	A self-collected EEG dataset comprises EEG recordings from 29 early PD patients, including 20 Women and 9 Men, aged between 52 and 74 years. These recordings were collected at the Department of Neurology, Tianjin Medical University General Hospital. EEG motions were noted at a sampling rate of 500 Hz. PD Oddball Dataset includes EEG recordings from PD patients and HC during a tristimulus auditory oddball paradigm. PD patients were assessed in two conditions: once while on medication and once after a 15-hour medication withdrawal. EEG signals were noted from 64 channels at a sampling rate of 500 Hz, filtered into five frequency bands, and analyzed for variances between PD and HC groups. The study features 24 pairs of matched subjects, ensuring balanced comparisons between the groups. Additionally, log spectral power features were extracted and mapped onto a brain topographic map to analyze neural biomarkers for Parkinson's disease progression.
L. Qiu, J. Li, L. Zhong, W. Feng [14]	The UC Dataset, collected by the University of California at San Diego, consists of at least 3 minutes of resting EEG data from 15 PD patients and 16 HCs, recorded using a 32-channel Biosemi Active Two EEG system at 512 Hz. The UNM Dataset, from the University of New Mexico, includes approximately 2 minutes of resting EEG data from 27 PD patients and 27 HCs, recorded with a 64-channel brain vision system at 500 Hz. The Iowa Dataset, from the University of Iowa, contains at least 2 minutes of resting EEG data from 14 PD patients and 14 HCs, also using a 64-channel brain vision system at 500 Hz. Qiu et al. utilized these three publicly available EEG datasets to validate the performance of their proposed MCPNet model for automatic PD detection.
E. Kumari, M. K. Shukla, O. Jee [15]	EEG brainwave feeling emotions dataset utilized for PD detection through EEG analysis. This dataset includes EEG recordings associated with 15 distinct emotions: shame, humiliation, contempt, disgust, fear, terror, enjoyment, joy, distress, anguish, surprise, anger, rage, interest, and excitement. The dataset was recorded for total 36 minutes from two adults, 1male and 1 female.

Machine learning and deep learning have markedly enhanced the detection of Parkinson's disease through speech analysis. Various algorithms, such as Support Vector Machines (SVM), Multi-Layer Perceptrons (MLP), and deep belief networks, have been utilized by researchers to analyze voice recordings of patients.

Studies indicate that SVM paired with Recursive Feature Elimination attained an accuracy of 93.84%, whereas MLP exhibited a notable accuracy of 98.31% in detecting PD from vocal features. These methodologies improve diagnostic accuracy and provide a cost-efficient solution for healthcare systems. The incorporation of feature selection methods and hyperparameter optimization enhances the efficacy of these models, establishing them as dependable instruments for the early detection and

monitoring of Parkinson's disease.

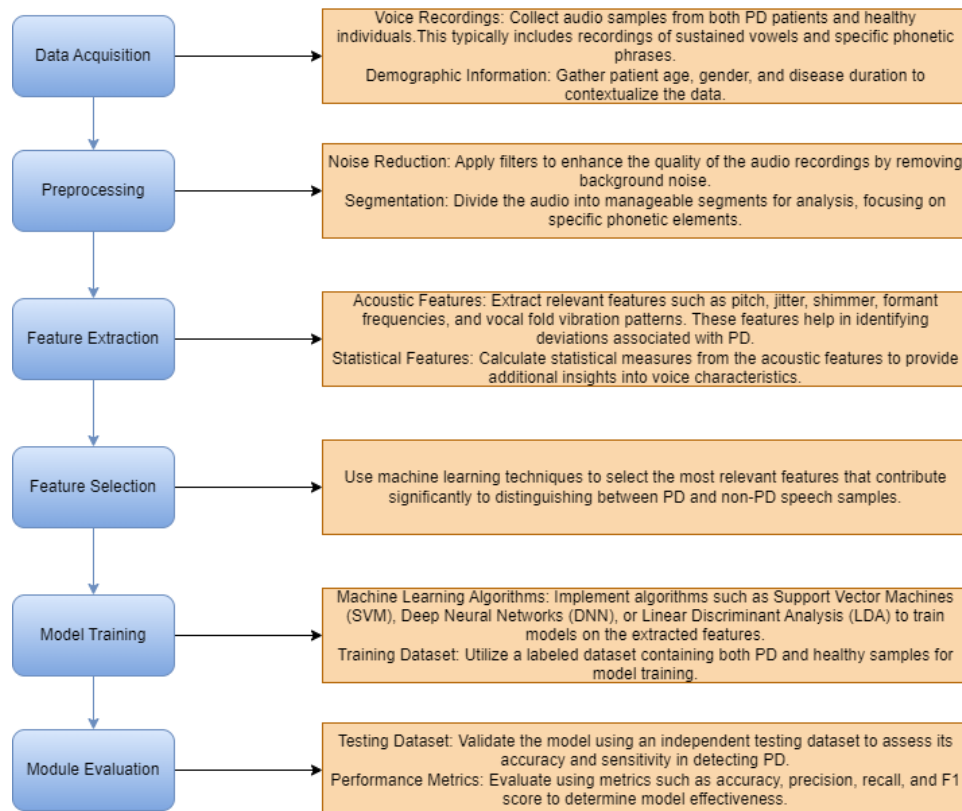


Figure 3 PD using Machine Learning and Speech Analysis

Figure 3 shows structured approach not only aids in understanding the process but also highlights critical areas where machine learning can significantly enhance early detection of Parkinson's disease through speech analysis.

Table 3 Speech Data Analysis

Authors	Work Done
A. Govindu and S. Palwe [16]	Multidimensional Voice Program (MDVP) dataset was collected from a cohort of 31 individuals with PD and the patients' ages ranged between 46 and 85 years. The normal readings were taken from participants who were aged 23 years. Everyone contributed an average of 6 phonation recordings and the dataset of 195 records were created per person. These records vary in duration, spanning from 1 to 36 seconds.
S. Grover, S. Bhartia, Akshama [17]	The Parkinson Telemonitoring Voice Data Set was utilized and gathered from the UCI Machine Learning Repository. The PD dataset utilized for early prediction of Parkinson's disease based on voice signals. It consists of 195 biomedical voice recordings; it is divided into 147 phonetics from PD cases and 48phonetics from healthy participants. The Parkinson's Telemonitoring Data Set contains biomedical voice recordings from 42 participant in the early stages of PD over a span of six months. Each of the 5,875 voice recordings is captured automatically by patients at home. It holds a unique data point used for predicting motor and total UPDRS scores. This dataset offers extensive data for detecting PD through voice analysis with 1040 instances and 27 attributes.
K. M. Alalayah, E. M. Senan [18]	The PD dataset comprises of 188 PD cases that includes 81 female participants and 107 male participants and 64 healthy samples in which 41 samples are females and 23 samples are males. Healthy participants were aged from 41 to 82 years, while ages of PD cases were ranged from 33 to 87 years. This PD speech dataset contains 753 features associated with speech characteristics.

M. T. García-Ordás, J. A. Benítez-Andrades [19]	The PC-GITA dataset includes 100 recordings of sustained vowels collected from native Colombian-Spanish speakers. It is divided into 50 participants with PD cases and the other 50 participants with healthy subjects. The SVD dataset consists of more than 2000 voice recordings which includes 71 different pathological conditions as well as recordings from healthy samples
---	--

Advanced models, such as Support Vector Machines (SVM) and artificial neural networks, have facilitated researchers in achieving high accuracy rates, with some models reaching up to 98.31% accuracy in classifying Parkinson's Disease (PD) based on vocal cues extracted from patient recordings. The approaches often include data pretreatment, feature selection, and hyperparameter optimization to enhance model performance. The incorporation of methodologies like the Synthetic Minority Over-sampling Technique (SMOTE) has proven effective in mitigating imbalanced datasets. These novel methodologies augment diagnostic accuracy and present possibilities for non-invasive screening techniques in clinical environments, perhaps enabling earlier identification and enhancing patient outcomes.

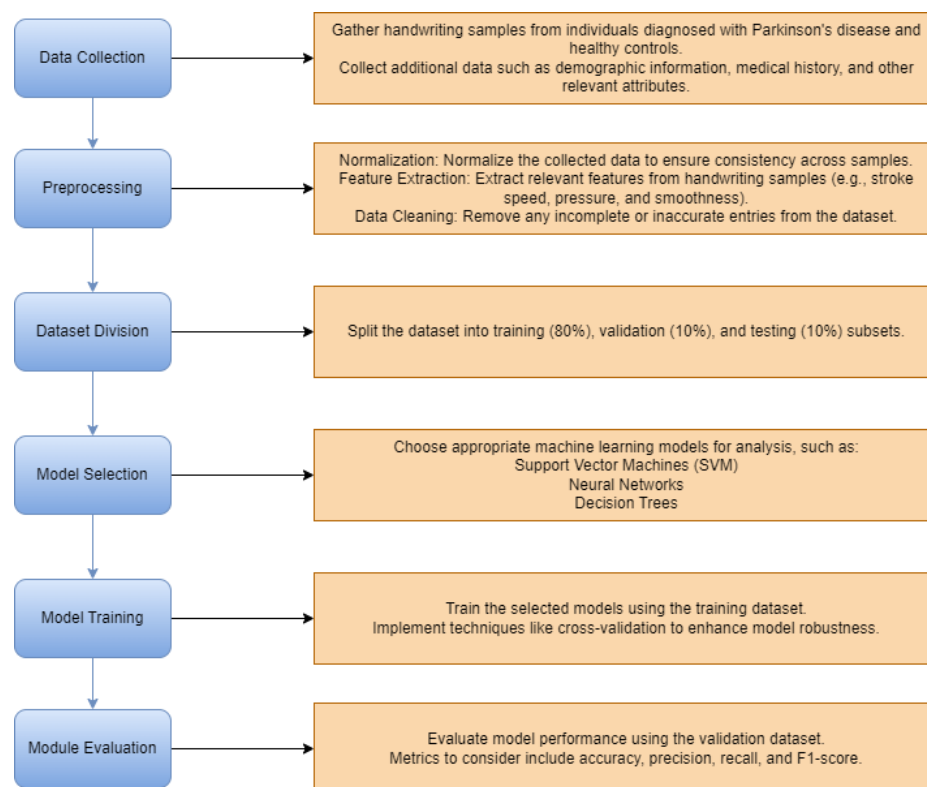


Figure 4 PD Detection using Handwriting Analysis and Machin Learning

Figure 4 shows approach highlights how handwriting analysis can be integrated into a machine learning framework for detecting Parkinson's disease effectively.

Table 4 Handwriting Dataset Analysis

Authors	Work Done
D. Impedovo, G. Pirlo, and G. Vessio [20]	Parkinson's disease handwriting (PaHaW) dataset includes writing activity performed by 75 right-handed subjects including 37 PD patients and 38 healthy controls, all native Czech speakers with at least ten years of education. The PD patients were assessed in their ON-state and evaluated for disease severity using the UPDRS score. Participants completed eight handwriting tasks recorded at a sampling rate of 200 Hz, capturing x- and y-coordinates, pressure, pen inclination, and button status.
M. Sultan, G. Hamed [21]	NewHandPD dataset includes handwriting samples from Parkinson's patients and healthy controls, featuring three types of handwriting images: meander, spiral, and circular. The entire data was divided into 60% for training and 40% for testing. Each handwriting type followed a specified workflow for model training. The dataset was used to

	evaluate various machine learning algorithms for accurate PD detection.
Z. Li, J. Yang, Y. Wang, M. Cai [22]	Online Arabic handwriting dataset includes samples from 30 Parkinson's disease patients (16 Female, 14 Male, average age 58) and 30 healthy controls (10 Female, 20 Male, average age 60). All subjects are native Arabic speakers, right-handed, with at least 6 years of schooling, and none had psychiatric symptoms or other central nervous system illnesses, except for Parkinson's disease. The dataset was collected at the Habib Bourguiba Hospital, Sfax, Tunisia, and includes data captured through a series of handwriting tasks.
P. Varalakshmi, B. T. Priya [23]	New - Offline hand drawings of Archimedes spirals dataset, includes samples from 43 PD patients and 43 age- and gender-matched HC, all right-handed and with no cognitive impairments. Each PD patient provided 3-5 spiral drawings, while each control provided one. The dataset was collected using a gel pen on a specified paper size without any drawing templates, preserving natural tremor and spacing characteristics. The participants were selected based on specific clinical criteria and excluded if they had conditions that could interfere with movement or cognitive function.
X. Wang, J. Huang, M. Chatzakou [24]	DraWritePD dataset consists of handwriting data from 20 PD patients and 29 healthy controls, captured using an iPad Pro during 12 tasks, recording features like coordinates, timestamp, azimuth, altitude, and pressure. NIATS dataset [133] comprises handwriting samples from 12 HC and 15 PD patients. Each participant provided three to four sine waves and one spiral drawing. The dataset contains two types of patterns: spiral and wave. For analysis, the dataset was divided into training and testing sets, with 72 images in the training set and 30 images in the testing set.

Complementary to Traditional Methods: Current diagnostic practices for PD typically involve a combination of patient medical history, symptom review, and neurological examinations⁴. Gait analysis can enhance these methods by providing objective data, particularly in assessing motor status and severity of symptoms.

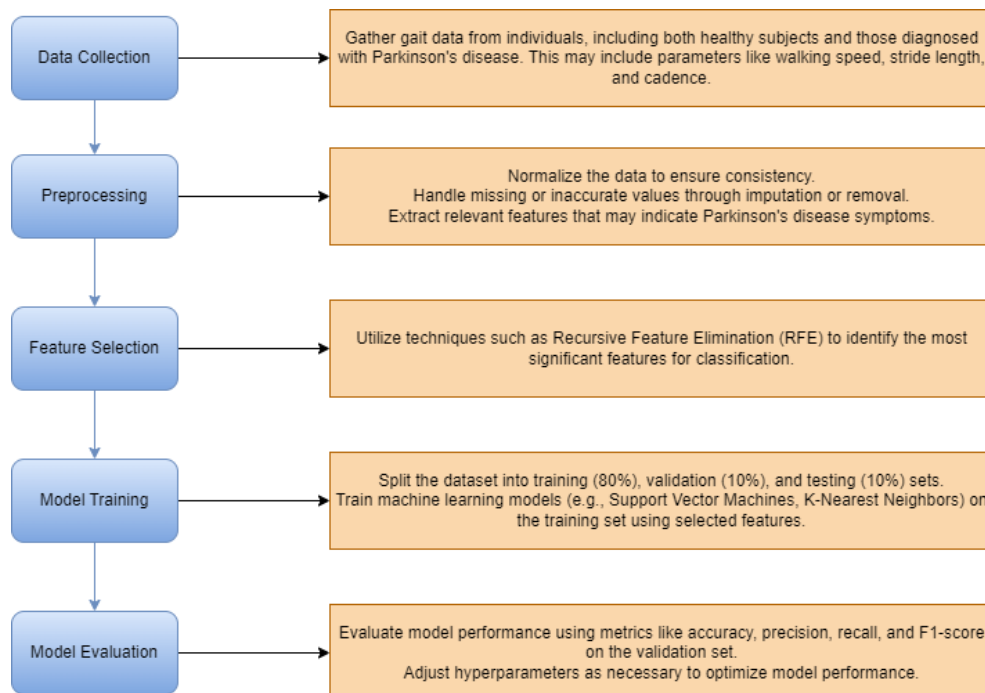


Figure 5 PD using Machine Learning and Gait Analysis

Figure 5 outlines a systematic approach to detecting Parkinson's disease through gait analysis, leveraging machine learning techniques for accurate diagnosis. Each step is crucial for ensuring that the model is trained effectively and can provide reliable predictions based on gait characteristics associated with Parkinson's disease

Table 5 Gait Data Analysis

Authors	Work Done
Q. Zeng, P. Liu, N. Yu, J. Wu, W. Huo [25]	The Gait based dataset is created from 54 participants with Parkinson's disease and 26 individuals without the condition and Gait scores were assigned to each participant. Healthy subjects were given a gait score of 0, while Parkinsonian gaits in the early stages were the focus, with all PD patients having UPDRS scores of 1 or 2. Video recordings were captured using two smartphones at a resolution of 720p and a frame rate of 25fps.
K. Hu, Z. Wang, S. Mei [26]	The vision-based dataset comprises videos obtained from 45 PD patients for gait analysis. A highly trained expert annotated the Freezing of Gait (FoG) ground truth for each video at the frame level. Videos were classified into FoG and non-FoG classes based on these annotations.
L. Wang, T. Tan, S. J. G. Lewis, and D. D. Feng [27]	utilized a dataset which contains videos from 45 subjects who underwent clinical assessment. The dataset consists of 167 videos, totalling approximately 25.5 hours of footage. These videos provide valuable data for studying and analysing different aspects of behaviour and movement in Parkinson's disease research.

Table 6 Hybrid Dataset based Analysis

Authors	Work Done
Z. Qin, Z. Jiang, J. Chen, C. Hu [28]	Qin et al. collected sEMG-based data from 147 PD patients, with UPDRS tremor severity levels (0-4) labeled by a neurosurgeon. From each patient's 10-second sEMG sequence (bicep, 1000 Hz), around seventy 2048-point sub-sequences were extracted. Totally, 10000 segments are gathered.
R. A. Zanini and E. L. Colombini [29]	A Standard 10-second electrocardiogram recordings from 60 subjects were analyzed, comprising 10 participants with established Parkinson's disease, 25 with early signs of Parkinson's disease, and 25 healthy controls. The recordings were analyzed to identify ECG patterns related to PD. The EMG Dataset includes 18 records with 116,000 data points each, from sessions with one ET patient and four PD patients. NinaPro provides EMG readings from 40 healthy subjects performing 49 actions, each repeated six times with rest intervals.
M. Y. Badr, R. A. Amer, M. A. Kotait [30]	The multi modal dataset contains 58 patients and 45 healthy samples. MRI scans and diffusion tensor imaging were captured from all the participants. They underwent extensive evaluations that encompassed detailed assessments in neurology, such as cognitive function and psychiatric status, comprehensive imaging techniques to examine brain structures and functions
Yang et al. [31]	Utilized data from the Parkinson's Progression Markers Initiative (PPMI) database, comprising 65 Parkinson's disease patients and 36 healthy controls. Each participant's data included both T1-weighted imaging (T1WI) and diffusion tensor imaging (DTI) for analysis.
G. Pahuja and B. Prasad [32]	The PPMI database includes 132 subjects in which 59 images are healthy controls and 73 images are PD patients. It focuses on those with complete datasets for neuroimaging, including MRI and SPECT SBR values, and biological markers such as CSF markers.
Boucherouite et al. [33]	Utilized a SPECT imaging model to categorize individuals as Healthy Controls using DICOM SPECT images from the PPMI database. The study used a dataset of 526 inputs, comprising 130 healthy and 396 images of affected individuals, to enhance the accuracy of early disease detection and differentiation from healthy controls.

3 Exploring Machine Learning and biochemical markers

Recent advancements in the identification of Parkinson's Disease (PD) have utilized machine learning (ML) and biochemical markers, with a particular emphasis on analysing voice signals. It has been demonstrated that ML techniques, such as Support Vector Machines (SVM) as well as Deep Learning (DL) models, can efficiently differentiate between healthy individuals and those with PD by analysing unique vocal features. Furthermore, a range of methodologies has been investigated, including the application of Synthetic Minority Over-sampling Technique (SMOTE) to tackle class imbalance in datasets, typically comprising voice recordings from PD patients.

Furthermore, the application of machine learning has been utilized to discover new subtypes of PD by

analyzing progressing disease rates and related genetic markers, which may lead to personalized treatment strategies. The combination of biochemical markers with voice analysis has the potential to enhance diagnostic capabilities, enabling non-invasive and reasonably priced screening methods for PD4. These innovative approaches highlight the significant impact of data mining in the early identification and treatment of Parkinson's Disease.

4 Conclusion

Research demonstrated that the analysis of micro-expressions revealed notable discrepancies in facial action units between Parkinson's disease patients and control subjects, indicating that smiling expressions were especially impacted by the condition. The amalgamation of facial expression analysis with machine learning not only improves diagnostic precision but also facilitates real-time observation of illness advancement. This technology may enable remote diagnosis, allowing healthcare personnel to monitor patient symptoms without requiring in-person consultations. The investigation of facial expression analysis as a diagnostic instrument for PD signifies a notable progress in neurology. Researchers are utilizing machine learning approaches to develop more efficient and non-invasive diagnostic tools that may boost patient outcomes and deepen our understanding of this complicated disease. The incorporation of machine learning methods in the detection of Parkinson's Disease has considerable potential for enhancing diagnostic accuracy and enabling early intervention. The integration of speech analysis, physiological indicators, and sophisticated computational models provides a comprehensive strategy for addressing the intricacies of Parkinson's disease diagnosis. As research advances, these approaches are anticipated to significantly improve patient outcomes and optimize diagnostic procedures in neurodegenerative disorders.

5 References

1. R. Souza, A. Winder, E. A. M. Stanley, V. Vigneshwaran, M. Camacho, R. Camicioli, O. Monchi, M. Wilms, and N. D. Forkert, "Identifying biases in a multicenter MRI database for Parkinson's disease classification: Is the disease classifier a secret site classifier?" *IEEE J. Biomed. Health Informat.*, vol. 28, no. 4, pp. 2047–2054, Apr. 2024, doi: 10.1109/JBHI.2024.3352513.
2. M. Camacho, M. Wilms, P. Mouches, H. Almgren, R. Souza, R. Camicioli, Z. Ismail, O. Monchi, and N. D. Forkert, "Explainable classification of Parkinson's disease using deep learning trained on a large multi-center database of T1-weighted MRI datasets," *NeuroImage, Clin.*, vol. 38, Jun. 2023, Art. no. 103405, doi: 10.1016/j.nicl.2023.103405.
3. S. Shinde, S. Prasad, Y. Saboo, R. Kaushick, J. Saini, P. K. Pal, and M. Ingahalikar, "Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI," *NeuroImage, Clin.*, vol. 22, Sep. 2019, Art. no. 101748, doi: 10.1016/j.nicl.2019.101748.
4. S. Chakraborty, S. Aich, and H.-C. Kim, "Detection of Parkinson's disease from 3T T1 weighted MRI scans using 3D convolutional neural network," *Diagnostics*, vol. 10, no. 6, p. 402, Jun. 2020, doi: 10.3390/diagnostics10060402.
5. V. Madge, V. S. Fonov, Y. Xiao, L. Zou, C. Jackson, R. B. Postuma, A. Dagher, E. A. Fon, and D. L. Collins, "A dataset of multicontrast unbiased average MRI templates of a Parkinson's disease population," *Data Brief*, vol. 48, Jun. 2023, Art. no. 109141, doi: 10.1016/j.dib.2023.109141.
6. T. M. Tassew, N. Xuan, and B. Chai, "PDDS: A software for the early diagnosis of Parkinson's disease from MRI and DaT scan images using detection and segmentation algorithms," *Biomed. Signal Process. Control*, vol. 86, Sep. 2023, Art. no. 105140, doi: 10.1016/j.bspc.2023.105140.
7. M. Ariz, R. C. Abad, G. Castellanos, M. Martínez, A. Muñoz-Barrutia, M. A. Fernández-Seara, P. Pastor, M. A. Pastor, and C. Ortiz-de-Solórzano, "Dynamic atlas-based segmentation and quantification of neuromelanin-rich brainstem structures in Parkinson disease," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 813–823, Mar. 2019, doi: 10.1109/TMI.2018.2872852.
8. R. Begum, T. P. Kumar, and M. R. N. Rao, "Deep convolutional neural networks for diagnosis of Parkinson's disease using MRI data," *Int. Inf. Eng. Technol. Assoc*
9. L. Qiu, J. Li, L. Zhong, W. Feng, C. Zhou, and J. Pan, "A novel EEGbased Parkinson's disease detection model using multiscale convolutional prototype networks," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–14, 2024, doi: 10.1109/TIM.2024.3351248.
10. E. Kumari, M. K. Shukla, O. Jee Pandey, and S. Yadav, "NeuroAid: Emotion-based EEG analysis for Parkinson's disease identification," *IEEE Sensors Lett.*, vol. 7, no. 12, pp. 1–4, Dec. 2023, doi: 10.1109/LENS.2023.3335226.
11. A. Govindu and S. Palwe, "Early detection of Parkinson's disease using machine learning," *Proc. Comput. Sci.*, vol. 218, pp. 249–261, 2023, doi: 10.1016/j.procs.2023.01.007.
12. S. Grover, S. Bhartia, Akshama, A. Yadav, and K. R. Seeja, "Predicting severity of Parkinson's disease using deep learning," *Proc. Comput. Sci.*, vol. 132, pp. 1788–1794, 2018, doi: 10.1016/j.procs.2018.05.154.
13. M. T. García-Ordás, J. A. Benítez-Andrades, J. Aveleira-Mata, J.-M. Alija-Pérez, and C. Benavides, "Determining the severity of Parkinson's disease in patients using a multi task neural

- network," *Multimedia Tools Appl.*, vol. 83, no. 2, pp. 6077–6092, Jan. 2024, doi: 10.1007/s11042-023-14932-x.
14. D. Impedovo, G. Pirlo, and G. Vessio, "Dynamic handwriting analysis for supporting earlier Parkinson's disease diagnosis," *Information*, vol. 9, no. 10, p. 247, Oct. 2018, doi: 10.3390/info9100247
 15. P. Varalakshmi, B. T. Priya, B. A. Rithiga, R. Bhuvaneaswari, and R. S. J. Sundar, "Diagnosis of Parkinson's disease from hand drawing utilizing hybrid models," *Parkinsonism Rel. Disorders*, vol. 105, pp. 24–31, Dec. 2022, doi: 10.1016/j.parkreldis.2022.10.020.
 16. M. Y. Badr, R. A. Amer, M. A. Kotait, S. M. Shoeib, and A. M. Reda, "Role of multimodal advanced biomarkers as potential predictors of cognitive and psychiatric aspects of Parkinson's disease," *Egyptian J. Neurol., Psychiatry Neurosurgery*, vol. 59, no. 1, p. 58, May 2023, doi: 10.1186/s41983-023-00662-2.
 17. G. Pahuja and B. Prasad, "Deep learning architectures for Parkinson's disease detection by using multi-modal features," *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105610, doi: 10.1016/j.combiomed.2022.105610.
 18. J. Boucherouite, A. Jilbab, and A. Jbari, "Automatic SPECT image processing for Parkinson's disease early detection," in *Smart Applications and Data Analysis SADASC (Communications in Computer and Information Science)*, vol. 1677, M. Hamlich, L. Bellatreche, A. Siadat, and S. Ventura, Eds., Cham, Switzerland: Springer, 2022, doi: 10.1007/978-3-031-20490-6_2.
 19. Pandey, P. (2025). Artificial intelligence for early detection of dental caries from radiographic images: A machine learning. *Journal of Dentistry and Clinical Sciences*, 1(1), 31–39. <https://agrpjournals.com/journals/journal/index.php/jdacs/article/view/12>
 20. Singh, A. R., & Kaur, A. (2025). AI adoption in agriculture: A managerial framework for sustainable agribusiness development. *Journal of Artificial Intelligence in Agriculture and Applied Research*, 1(1), 1–12. <https://agrpjournals.com/journals/journal/index.php/jaiaar/article/view/2>