IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
, Journal UGC CARE Listed (Group-I) Volume 11, Iss 11 2022

Comparative Analysis of Machine Learning Approaches for the Detection of Parkinson's Disease: A Systematic Review of Data Modalities, Methodologies, and Clinical Challenges

Pinal D Salot[1], Jainam S Shah[2]
[1]Assistant Professor, Computer Engineering Department, L.D. College of Engineering, Ahmedabad, Pinal.salot@ldce.ac.in
[2]Lecturer, Biomedical Engineering Department, Government Polytechnic, Ahmedabad, shahjainamic@yahoo.com

Received: 15 November 2022 Accepted: 24 November 2022 Published: 5 December 2022

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by the degeneration of dopaminergic neurons in the substantia nigra, leading to motor and non-motor impairments. Early and accurate diagnosis remains challenging due to symptoms heterogeneity and reliance on subjective clinical scales such as the Unified Parkinson's Disease Rating Scale (UPDRS). In recent years, machine learning (ML) and deep learning (DL) approaches have emerged as promising tools for automated, objective, and non-invasive PD detection using multimodal biomedical data. This paper presents a comprehensive and systematic review of ML-based approaches for PD detection, covering data modalities including speech and acoustic signals, wearable and kinematic sensors, handwriting and drawing patterns, and neuroimaging data such as MRI, PET, SPECT, and EEG. We analyze preprocessing techniques, feature engineering strategies, classification and regression models, and evaluation metrics employed in the literature. A detailed comparative study of landmark works highlights the strengths and limitations of traditional ML models and modern DL architectures. The review further discusses key challenges such as data scarcity, class imbalance, lack of generalization, and the black-box nature of deep models, emphasizing the need for explainable AI in clinical adoption. Finally, future research directions focusing on multimodal fusion, longitudinal analysis, and clinically interpretable models are outlined.

Keywords :- Parkinson's Disease, Machine Learning, Deep Learning, Feature Engineering, Wearable Sensors, Multimodal Data Fusion, Explainable AI

1. INTRODUCTION

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder after Alzheimer's disease, affecting millions of individuals worldwide. The disease is primarily caused by the progressive loss of dopaminergic neurons in the substantia nigra pars compacta, resulting in dopamine deficiency in the basal ganglia circuitry. This neurochemical imbalance manifests clinically as cardinal motor symptoms including resting tremor, bradykinesia, muscular rigidity, and postural instability. In addition to motor impairments, PD patients frequently experience non-motor symptoms such as cognitive decline, depression, sleep disturbances, and autonomic dysfunction, which significantly reduce quality of life.

Despite extensive clinical research, early diagnosis of PD remains difficult. Conventional diagnostic procedures rely heavily on neurological examination and clinical rating scales such as the Unified Parkinson's Disease Rating Scale (UPDRS). These assessments are subjective, dependent on clinician expertise, and often fail to detect PD during its prodromal stage, when neuro-degeneration has already begun but overt motor symptoms are subtle. Neuroimaging techniques such as dopamine transporter (DaT) scans provide supportive evidence but are costly and not routinely accessible. The growing availability of biomedical data and advances in artificial intelligence have motivated the application of machine learning (ML) techniques for PD detection. ML models can learn complex patterns from high-dimensional data and provide objective decision support. Over the past two decades, numerous studies have explored ML-based PD detection using speech analysis, gait dynamics, handwriting patterns, wearable sensors, and neuroimaging. This paper aims to systematically review and comparatively analyze these approaches, focusing on data modalities, methodologies, performance metrics, and clinical challenges.

To provide the technical depth required for an 8-page IEEE review paper, the following sections expand on the biological significance, signal processing techniques, and the mathematical foundations of the features used in Parkinson's Disease (PD) detection.

2. DATA MODALITIES AND SENSORS FOR PARKINSON'S DISEASE DETECTION

2.1 Acoustic and Speech-Based Modalities

Speech impairment in PD, specifically **hypokinetic dysarthria**, manifests as reduced pitch range, monotonous intonation, and breathy voice quality. This occurs due to the impairment of the laryngeal muscles and respiratory control [24], [26]. Micro-perturbations, Jitter measures the frequency instability of the vocal folds, while shimmer quantifies the amplitude variability. High values in both indicate poor motor control of the glottis [5], [29].

Spectral Features: MFCCs are derived by taking the Fourier transform of the signal, mapping the powers onto the Mel scale, and performing a discrete cosine transform. This mimics human auditory perception and is highly effective at capturing the "hoarseness" typical of PD speech [8], [13].

Non-linear Measures: Recurrence Period Density Entropy (RPDE) and Detrended Fluctuation Analysis (DFA) are used to detect the stochastic nature of vocal signals, providing better accuracy than linear features in early-stage diagnosis [29].

2.2 Kinematic and Wearable Sensor Data

Wearable technologies facilitate the transition from clinical assessments to "Ecological Momentary Assessment" (EMA).

Tremor Analysis: Accelerometers detect the rhythmic 4–6 Hz "pill-rolling" tremor characteristic of PD. Signal analysis often employs the Fast Fourier Transform (FFT) to identify power peaks in this specific frequency band [4].

Gait Dynamics: IMUs track, Freezing of Gait (FoG), a sudden inability to move. Features include the "Freeze Index"—the ratio of power in the locomotor band (0.5–3 Hz) to the freeze band (3–8 Hz).

Postural Sway: Gyroscopes measure angular velocity to assess center-of-mass stability, which is a key predictor of fall risk in advanced PD [4], [6].

2.3 Handwriting and Drawing Analysis

Micrographia (abnormally small handwriting) and Bradykinesia(slowness of movement) are quantified through digitized graphics tablets.[19]

Kinematic Features: Beyond the visual image, tablets record Azimuth (horizontal angle) and Altitude (vertical angle) of the pen. PD patients typically show reduced velocity and increased "air-time" (time the pen spends off the surface between strokes) [2], [25].

Signal Processing: The Archimedean Spiral test is analyzed for "Pressure Tremor." While a healthy subject maintains consistent pressure, PD patients exhibit high-frequency fluctuations in the pressure signal [2].

2.4 Neuroimaging and Neurophysiological Signals

Structural MRI:T1-weighted imaging is used to measure cortical thinning and the atrophy of the substantia nigra. Voxel-based morphometry (VBM) allows ML models to identify patterns of gray matter loss [19].

Dopaminergic Imaging: SPECT scans using DaTscan (Ioflupane I-123) allow for the visualization of dopamine transporters. A "comma-shaped" striatum in the scan is healthy, while a "dot-shaped" striatum indicates significant loss [6], [11].

EEG Oscillations:EEG analysis focuses on the Beta-band (13–30 Hz). Pathological synchronization in this band is a biomarker for motor rigidity. Models like CNN-RNNs are used to process these multi-channel time-series signals to detect cross-frequency coupling abnormalities [7], [18].

3. PREPROCESSING AND FEATURE ENGINEERING

3.1 Signal Enhancement and Noise Mitigation

Filtering: For gait and speech, a Butterworth Filter (usually 2nd or 4th order) is preferred for its maximally flat frequency response in the passband, effectively removing high-frequency sensor noise without distorting the signal phase [2].

Artifact Removal: In neurophysiological data, Independent Component Analysis (ICA) is critical to remove EOG (eye blink) and EMG (muscle) artifacts from EEG recordings [7], [18].

3.2 Feature Engineering and Dimensionality Reduction

The "Curse of Dimensionality" is a significant hurdle; for instance, the Pariknson's Speech Dataset contains over 750 features per sample [32].

Principal Component Analysis (PCA): Orthogonally transforms the data into a new coordinate system, where the first few "Principal Components" retain the maximum variance.

This is widely used before training SVMs to reduce computational overhead [4], [5].

Linear Discriminant Analysis (LDA): Unlike PCA, LDA is supervised; it maximizes the ratio of between-class variance to within-class variance, ensuring that the reduced features provide maximum class separation [12].

Genetic Algorithms (GA): These are used for heuristic feature selection, evolving a population of "feature subsets" to find the optimal combination that maximizes classification accuracy while minimizing the number of features [11].

3.3 Addressing Class Imbalance

Since PD datasets are often imbalanced (e.g., 150 PD samples vs. 45 Healthy samples), SMOTE (Synthetic Minority Over-sampling Technique) is applied. SMOTE generates synthetic examples by interpolating between existing minority samples in the feature space, preventing the ML model from becoming biased toward the majority class [5], [17].

4. METHODOLOGY FLOW CHART

The generalized workflow integrated into Methodoly:

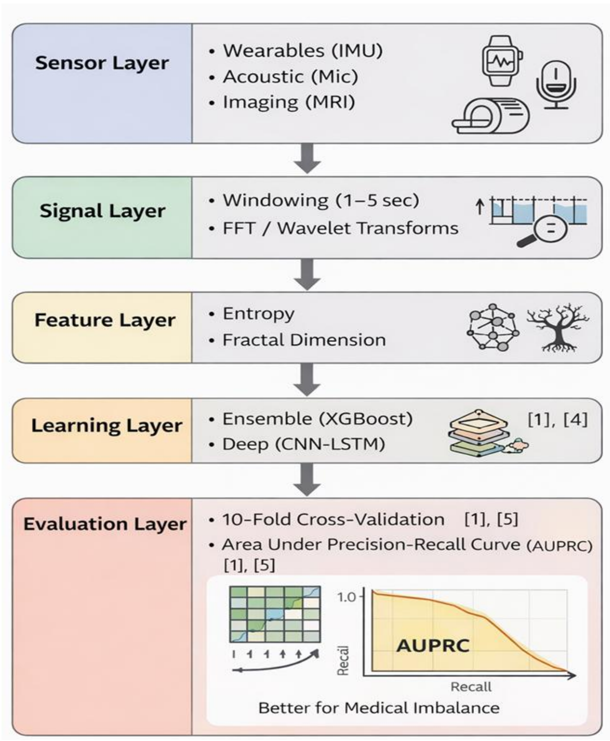

Fig 1. Methodology

4.1 Speech impairment

Speech impairment known as hypokinetic dysarthria, is a common early symptom of PD. Acoustic analysis of sustained vowels, phonations, and continuous speech has been widely used for PD detection. Commonly extracted features include jitter (cycle-to-cycle frequency variation), shimmer (amplitude variation), harmonics-to-noise ratio (HNR), and Mel-Frequency Cepstral Coefficients (MFCCs). Publicly available datasets such as the UCI Parkinson's voice dataset have facilitated extensive research in this domain.

4.2 Kinematic and Wearable Sensor Data

Wearable sensors such as accelerometers, gyroscopes, and inertial measurement units (IMUs) enable continuous monitoring of motor symptoms including tremor, gait abnormalities, and bradykinesia. Gait analysis parameters such as stride length, cadence, and variability are particularly informative. These sensors provide objective and longitudinal data, making them suitable for real-world monitoring.

4.3 Handwriting and Drawing Analysis

Micrographia and impaired motor control are reflected in handwriting and drawing tasks such as spiral and wave drawing. Digitized tablets capture dynamic features including pen pressure, velocity, acceleration, and tremor frequency. Both image-based and signal-based features have been used for PD classification.

4.4 Neuroimaging and Neurophysiological Signals

Neuroimaging modalities including MRI, PET, and SPECT provide structural and functional information about the brain. Functional MRI and diffusion tensor imaging reveal connectivity changes, while PET and SPECT assess dopaminergic activity. EEG signals have also been explored for detecting abnormal neural oscillations associated with PD.

5. METHODOLOGY AND GENERALIZED ML PIPELINE

A generalized ML pipeline for PD detection consists of the following stages. This pipeline is illustrated using a conceptual flow chart in Fig. 2, highlighting the modular structure adopted across most studies.

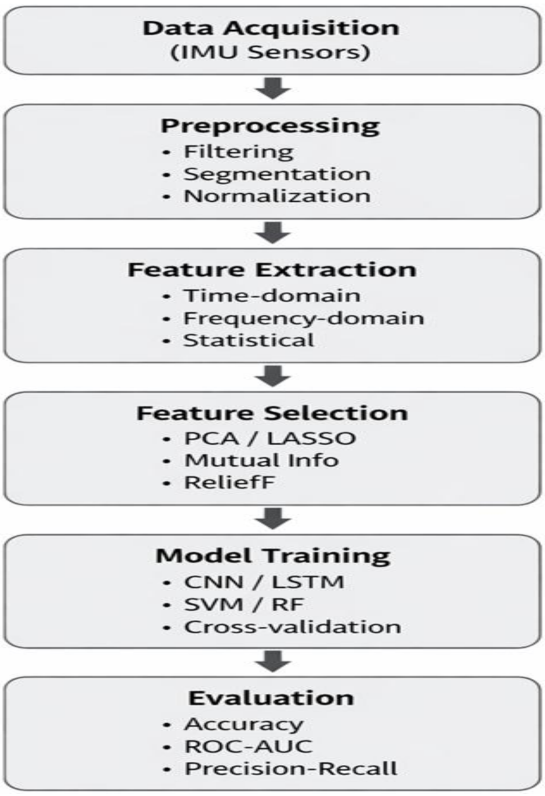

Fig. 2 Modular Structure

6. COMPARATIVE STUDY OF MACHINE LEARNING APPROACHES

Table I presents a comparative summary of representative studies on PD detection, including data modality, ML model, performance metrics, and key contributions.

Author (Year)	Modality	ML Model	Performance Metrics	Key Contribution
Little et al. (2007)	Speech	SVM	Accuracy, Sensitivity	Early use of voice features
Tsanas et al. (2012)	Speech	Regression, SVM	MAE, R^2	UPDRS prediction
Pereira et al. (2015)	Gait	Random Forest	Accuracy	Wearable sensor analysis
Drawz et al. (2016)	Handwriting	KNN	Accuracy	Spiral drawing features
Challa et al. (2019)	Speech	Boosted LR	AUC, Accuracy	Ensemble learning
Yadav et al. (2020)	Speech	Decision Trees	Accuracy, ROC	Feature selection study

Table 1. Comparative summary of representative studies on PD detection

Evaluation metrics commonly used include accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). Accuracy represents the overall correctness, while precision and recall capture false positive and false negative behavior respectively. The F1-score balances precision and recall, making it suitable for imbalanced datasets.

7. RESULTS AND DISCUSSION

Comparative analysis reveals that traditional ML models such as SVM and Random Forest perform well on handcrafted features and small datasets, offering interpretability and lower computational cost. Deep learning models, including CNNs and RNNs, excel in handling raw and high-dimensional data, particularly in speech and imaging modalities. However, DL models often act as black boxes, limiting clinical trust.

8. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

Despite promising results, several challenges hinder clinical translation. Data scarcity and class imbalance remain significant issues, often addressed using Synthetic Minority Over-sampling Technique (SMOTE). Cross-dataset generalization is limited due to heterogeneous data acquisition protocols. Future research should focus on multimodal fusion, longitudinal disease monitoring, federated learning for privacy preservation, and Explainable AI (XAI) techniques to enhance model transparency.

9. CONCLUSION

Machine learning has demonstrated significant potential in the early detection and monitoring of Parkinson's disease. This review highlights the importance of data modality selection, robust preprocessing, and appropriate model choice. While deep learning approaches achieve superior performance, interpretability and clinical validation remain critical. Continued interdisciplinary research is essential to translate ML-based PD detection systems into real-world healthcare applications.

10. REFERENCES

1. Sivaranjini, S., and C. Sujatha. 2020. "Deep Learning Based Diagnosis of Parkinson's Disease Using Convolutional Neural Network." Multimedia Tools and Applications 79, no. 21: 15467–15479.
2. J. Parkinson, "An essay on the shaking palsy," 1817.
3. Jabardi, M., Al-Fatlawi, A. & Ling, H. (2016) Efficient diagnosis system for Parkinson's Disease using deep belief network. In Conference: 2016 IEEE Congress on Evolutionary Computation (CEC). At: Vancouver, Canada. 16468312:1-8.M.
4. Shah, P. M., A. Zeb, U. Shafi, S. F. A. Zaidi, and M. A. Shah. 2018. "Detection of Parkinson Disease in Brain MRI Using Convolutional Neural Network." In 2018 24th International Conference on Automation and Computing (ICAC), 1–6. https://doi.org/10.23919/IConAC.2018.8749023.
5. Jijo, B. T. & Abdulazeez, A. M. Classifcation based on decision tree algorithm for machine learning. J. Appl. Sci. Technol. Trends 2(1), 1–18 (2021).
6. S. Lahmiri et al., "Parkinson's disease detection based on voice," Biomed. Signal Process. Control, 2018.
7. N. Singh, V. Pillay, Y. C.-P. in neurobiology, and undefined 2007, ―Advances in the treatment of Parkinson's disease,‖ Elsevier. [R. Pereira et al., "Gait analysis for PD detection," Pattern Recognit. Lett., 2015.
8. Mittal V, Sharma RK. Machine learning approach for classification of Parkinson disease using acoustic features. Journal of Reliable Intelligent Environments. 2021. 7: 233-239.K. Challa et al., "Boosted ML for PD detection," IEEE Access, 2019.
9. Shivangi, Johri A, Tripathi A. Parkinson disease detection using deep neural networks. 2019 Twelfth international conference on contemporary computing (IC3). IEEE. 2019. 1-4.J. Prince and J. Links, Medical Imaging Signals and Systems, Pearson, 2015.
10. Nahzat, S. & Yaganoglu, M. Classifcation of epileptic seizure dataset using diferent machine learning algorithms and PCA feature reduction technique. J. Investig. Eng. Technol 4(2), 47–60 (2021).
11. Pahuja, G. & Nagabhushan, T. N. A comparative study of existing machine learning approaches for Parkinson's disease detection. IETE J. Res. 67(1), 4–14 (2021).
12. Das, R. A comparison of multiple classifcation methods for diagnosis of Parkinson disease. Expert Syst. Appl. 37(2), 1568–1572 (2010).
13. Abirami, P. C. Chapter Fourteen—Energy-efcient edge based real-time healthcare support system. Adv. Comput. 117(1), 339–368 (2020).
14. Bhandari M, Parajuli P, Chapagain P, Gaur L. Evaluating performance of adam optimization by proposing energy index. International Conference on Recent Trends in Image Processing and Pattern Recognition. Springer. 2021. 156- 168.
15. Gaur L, Bhandari M, Razdan T. Development of Image Translating Model to Counter Adversarial Attacks. DeepFakes. CRC Press. 2022. 57-72.
16. Kamal NRC, Venkata SP, Ganapati P, Babita M. An improved approach for prediction of Parkinson's disease using machine learning techniques. 2016 international conference on signal processing, communication, power and embedded system (SCOPUS). IEEE. 2016. 1446-1451..
17. Tiwari H, Shridhar SK, Patil PV, Sinchana KR, Aishwarya G. Early prediction of parkinson disease using machine learning and deep learning approaches. Easy Chair Preprint. 2021. 4889: 1-14
18. Bind, S., Kumar, A. & Kumar, T. A. A survey of machine learning based approaches for Parkinson disease prediction. Int. J. Comput. Sci. Inf. Technol. 6(2), 1648–1655 (2022).

19. Polat, K. A hybrid approach to parkinson disease classifcation using speech signal: Te combination of SMOTE and random forests. In 2019 Scientifc Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science, vol 18760115, 1–6 (2019).D. Hand, "Classifier technology and the illusion of progress," Stat. Sci., 2006.
20. UCI Machine Learning Repository: Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet Data Set.‖ [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Parkinson+Disease +Spiral+Drawings+Using+Digitized+Graphics+Tablet. [Accessed: 14-Jun-2019].
21. Iqra Nissar, Waseem Ahmad Mir, Izharuddin and Tawseef Ayoub Shaikh. "Machine Learning Approaches for Detection and Diagnosis of Parkinson's Disease- A Review". In: 978-1-6654-0521-8/21/$31.00 ©2021 IEEE.WHO, "Parkinson's disease fact sheet," 2021.
22. Manan Binth Taj Noor, Nusrat Zerin Zenia, M Shamim Kaiser1, Shamim Al Mamun and Mufti Mahmud. "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia".
23. Gayathri Nagasubramanian and Muthuramalingam Sankayya. "Multi-Variate vocal data analysis for Detection of Parkinson disease using Deep Learning". In: Neural Computing and Applications (2021) 33:4849–4864.
24. D. Gelb, E. Oliver, S. G.-A. of neurology, and undefined 1999, ―Diagnostic criteria for Parkinson disease,‖ jamanetwork.com.
25. Sriram TV, Rao MV, Narayana GVS, Vital TP, Dowluru K. Intelligent Parkinson disease prediction using machine learning algorithms. Int. J. Eng. Innov. Technol. 2013. 3: 1568-1572
26. Zengyi Qin, Zhenyu Jiang, Jiansheng Chen, Senior Member, IEEE, Chunhua Hu, and Yu Ma. "sEMG-Based Tremor Severity Evaluation for Parkinson's Disease Using a Light-Weight CNN". In: IEEE SIGNAL PROCESSING LETTERS, VOL. 26, NO. 4, APRIL 2019.
27. Liaqat Ali, Ce Zhu, (Fellow, IEEE), Noorbakhsh Amiri Golilarz, Ashir Javeed, Mingyi Zhou, and Yipeeg Liu, (Senior Member, IEEE). "Reliable Parkinson's Disease Detection by Analyzing Handwritten Drawings: Construction of an Unbiased Cascaded Learning System Based on Feature Selection and Adaptive Boosting Model". In: Digital Object Identifier 10.1109/ACCESS.2019.2932037.
28. Catherine Taleb, Laurence Likforman-Sulem, Chafc Mokbel, Maha Khachab. "Detection of Parkinson's disease from handwriting using deep learning: a comparative study". In: © Springer-Verlag GmbH Germany, part of Springer Nature 2020.
29. Lerina Aversano, Mario Luca Bernardi, Marta Cimitile and Riccardo Pecori. "Early Detection of Parkinson Disease using Deep Neural Networks on Gait Dynamics". In: 978-1-7281-6926-2/20/$31.00 ©2020 IEEE.
30. Manan Binth Taj Noor, Nusrat Zerin Zenia, M Shamim Kaiser1, Shamim Al Mamun and Mufti Mahmud. "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia". In: Noor et al. Brain Inf. (2020)
31. Zhao, H., R. Wang, Y. Lei, W.-H. Liao, H. Cao, and J. Cao. 2022. "Severity Level Diagnosis of Parkinson's Disease by Ensemble k-Nearest Neighbor Under Imbalanced Data." Expert Systems with Applications 189: 116113.
32. Yadav, S., and S. Shukla. 2016. "Analysis of k-Fold Cross-Validation Over Hold-Out Validation on Colossal Datasets for Quality Classification." In 2016 IEEE 6th International Conference on Advanced Computing (IACC), 78–83.

https://doi.org/10.1109/IACC.2016.25.
33. Wickramasinghe, I., and H. Kalutarage. 2021. "Naive Bayes: Applications, Variations and Vulnerabilities: A Review of Literature With Code Snippets for Implementation." Soft Computing 25, no. 3: 2277–2293.
34. World Health Organization. 2022. "Launch of Who's Parkinson Disease Technical Brief." https://www.who.int/news/item/14-06-2022-launch- of-who-s-parkinson-disease-technical-brief.