

Review on Harnessing Artificial Intelligence for Parkinson's Disease

Pragadeeswaran Sundaresan¹, Raju T², Tamilselvan T³, Thiyaagu T⁴, Govinda Rajulu. G⁵ and Ganeshan R⁶
^{1,3-6}School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, India

²Department of CSE (Cyber Security), 2CMR College of Engineering & Technology, Kandlakoya, Medchal, Hyderabad, India

Email: pragadeesresearch@gmail.com, t.raju@cmrcet.ac.in, tamilselvan.akash@gmail.com, thiyagu.57@gmail.com, rajulug7@gmail.com, ganeshramasamy111@gmail.com

Abstract— Parkinson's disease (PD) is a long-lasting and worsening movement disease that is mostly caused by the death of neurons in the substantia nigra, a part of the brain that is very important for controlling movement. This neurodegenerative disorder presents with motor symptoms including tremors, bradykinesia (slowness of movement), stiffness, and postural instability. Along with these apparent indicators, many patients have non-motor symptoms that make their quality of life much worse. These include depression, cognitive impairment, sleep problems, and autonomic dysfunction. PD is the second most common neurodegenerative illness in the world, behind Alzheimer's. It affects more than 10 million individuals. As the population ages, the number of people with PD is likely to climb sharply, making it more important than ever to have a quick and precise diagnosis. However, identifying PD in its early stages is still very hard because its symptoms are similar to those of other neurological illnesses and there are no clear biomarkers.

Keywords— UCI PD Dataset, PPMI (Parkinson's Progression Markers Initiative), Power NewHandPD, PC-GITA.

I. INTRODUCTION

Artificial Intelligence (AI) has become a strong partner in dealing with this problem. AI models can find hidden patterns that may come before clinical symptoms by looking at complicated, high-dimensional data like brain imaging, speech patterns, handwriting dynamics, and gait metrics. Machine Learning (ML) and Deep Learning (DL) approaches, in particular, may learn from past patient data and provide very accurate predictions about when and how PD will start and get worse. This new technology has the potential to change the way PD is diagnosed, monitored, and treated, making it possible to create tailored and proactive healthcare plans.

II. THE NEED FOR AI IN PD DIAGNOSIS

Traditional diagnostic methods depend on clinical symptoms and subjective evaluations, frequently leading to delayed or erroneous diagnoses. AI models can find hidden signs in voice, handwriting, gait, and neuroimaging data. These tools give you diagnostic predictions that are consistent, scalable, and often more accurate.

A. Accessibility of PD Datasets

We begin by exploring the availability and characteristics of publicly accessible PD datasets. These datasets are foundational for training and validating ML and DL models, and we examine how researchers have leveraged them in recent studies.

B. Preprocessing of Multimodal Data

Next, we look at the different ways that data preprocessing was done on clinical, voice, handwriting, and imaging data. This part talks about how techniques like normalization, filling in missing values, and scaling features can make input data better and more consistent, which greatly improves the performance of models.

C. Machine Learning and Deep Learning Applications

The review's main focus is on the use of supervised and unsupervised machine learning (ML) algorithms, as well as more advanced deep learning (DL) models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs).

We look at how these methods are used on clinical records, neuroimaging data, genetic profiles, and sensor data from wearable devices.

D. Trends and Innovations in DL Architecture

We find new trends and cutting-edge methods that are changing the future of AI-based PD diagnosis by looking at deep learning architectures, training strategies, and optimization techniques.

E. Critical Evaluation of Existing Models

We critically evaluate contemporary ML and DL models regarding their accuracy, robustness, and practical applicability. The review talks about the pros and cons of different methods, as well as the technical and moral problems that come up when using them in clinical settings.

F. Systematic Review Structure

The framework of this review is structured. After surveying the variety of PD-related datasets, it explores preprocessing techniques and then takes a close look at ML and DL frameworks for illness diagnosis.

G. Insights and Future Directions

Finally, we put all of our findings together to show where there are gaps in knowledge, suggest new areas of study, and give practical advice on how to use AI to improve PD detection. The research finishes by summarizing the implications of the studied literature for building more accurate, accessible, and trustworthy diagnostic tools employing machine and deep learning technologies.

III. DATA MODALITIES USED IN AI RESEARCH

AI approaches use several data types:

- Voice recordings (vocal tremors, pitch changes)
- Handwritten spiral drawings
- Gait data using wearable sensors
- Brain scans (MRI, EEG)
- Smartphone and smartwatch data

IV. PUBLIC DATASETS FOR PARKINSON'S DISEASE

TABLE I: KEY DATASETS AVAILABLE FOR AI-BASED PD RESEARCH

Dataset Name	Data Type	Details
UCI PD Dataset	Voice	Biomedical voice signals from 31 patients
PPMI	MRI/Clinical	Longitudinal imaging and clinical metrics
NewHandPD	Handwriting	Spiral/meander drawings from 92 participants
mPower	Mobile	Sensor data from tapping/walking tests
PC-GITA	Speech	Spanish dataset with vocal biomarkers

V. DATA PREPROCESSING TECHNIQUES

Data preprocessing plays a crucial role in enhancing the accuracy of AI models. Common techniques include:

TABLE II: KEY DATASETS AVAILABLE FOR AI-BASED PD RESEARCH.

Dataset	Description	Data modality	References
Parkinsons[10]	197 biomedical voice recordings in CSV format, collected from 31 individuals, including 23 with PD, by Max Little and the National Centre for Voice and Speech.	CSV	[24–27]
Italian Parkinson's voice and speech database [12]	Audio recordings from 37 healthy participants and 28 PD patients, collected by Giovanni Dimauro et al. [11].	WAV & XLSX	[28]
Parkinson speech dataset with multiple types of sound recordings[13]	1040 sound recordings from 40 individuals, including 20 healthy and 20 with PD, collected at Istanbul University.	WAV	[28, 30]
Parkinsons telemonitoring[14]	Biomedical voice measurements from 42 individuals with early-stage PD, with 5,875 recordings in ASCII format.	ASCII	[31]
Parkinson's disease classification[15]	756 voice recordings of 188 PD patients and 64 healthy subjects from Istanbul University.	CSV	[32]
PPMI[19]	A longitudinal study providing clinical, imaging, and biospecimen data accessible after registration.	Various	[33]
mPower[21]	3711 records of tapping activity and data from memory, walking, and voice activities.	Various	[31]
MDVR-KCL dataset[22]	Voice recordings from 37 participants, including 16 PD patients and 21 healthy controls.	WAV	[34, 35]
Gaitin Parkinson's disease [23]	279 gait measures from 93 patients with idiopathic PD and 73 healthy controls, capturing walking patterns.	CSV	[36]
Parkinson disease spiral drawings using digitized graphics tablet [16]	Records from 62 PD patients and 15 healthy individuals, with handwriting assessments using a graphics tablet.	Digital images	[16]
Early biomarkers of Parkinson's disease based on natural connected speech [17]	Dataset includes demographic and clinical information on 30 patients with early-stage PD and 50 healthy controls.	XLS & CSV	[17]

- Normalization and standardization
- Dimensionality reduction (PCA)
- Noise filtering and smoothing
- Spectrogram transformations (for voice data)
- Data augmentation using GANs

VI. PERFORMANCE OF ML MODELS

TABLE III: SUMMARY OF ML MODEL PERFORMANCES ON PD DATASETS

Model	Accuracy	Precision	Recall	F1 Score
XGBoost	98.75%	96.50%	93.41%	96.73%
SVM	90.01%	-	89.01%	-
KNN	90.20%	90.01%	89.30%	89.50%

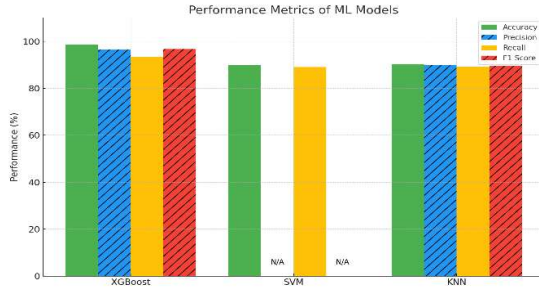


Fig 1. Performance Metrics of ML Models

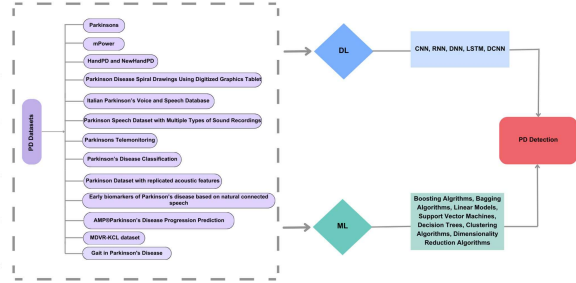


Fig 2. Taxonomy of ML and DL approaches for PD detection

Taxonomy of AI Approaches

VIII. DEEP LEARNING METHODS AND RESULTS

DL models like CNN, LSTM, and hybrid CNN-LSTM architectures are ideal for speech and imaging tasks. EfficientNet-V2 and AlexNet have achieved over 99% accuracy in classifying MRI and voice signals. These models automatically learn discriminative features, improving generalization across datasets.

TABLE IV: PERFORMANCE OF ML MODELS FOR PD DETECTION

References	Dataset	Model	Acc	Precision	F1-score	Recall	Specificity	AUC	MSE
[56]	NewHandPD [52]	SVM	90.01%	—	—	89.01%	93.01%	—	—
[62]	Speech dataset from UCI [65]	XGBoost	98.75%	96.50%	96.73%	93.41%	—	93.33%	—
[23]	Parkinsons [10]	EESDPD	93.20%	93.75%	95.60%	97.6%	—	—	—
[63]	PPMI [19]	PLR	89.30%	—	—	—	—	93.90%	—
[64]	UCI [65]	Extra Tree Classifier	92.56%	95.89%	94.95%	94.56%	—	97.78%	—
[66]	PPMI [19]	ANN	—	—	—	90.01%	56.01%	—	—
[67]	mPower[21]	RWSL	—	—	—	—	—	—	2.76%
[26]	Parkinsons [10]	MASS-PCNN	91.10%	97.80%	99.50%	94.70%	—	—	—
[27]	Parkinsons [10]	KNN	90.20%	90.01%	89.50%	89.30%	—	—	—
[69]	PPMI [19]	SVM, MFW	98.30%	—	—	98.51%	97.90%	—	—

Accuracy Distribution Among ML Models for PD Detection

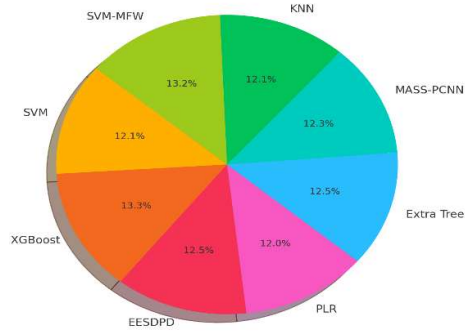


Fig 3. Accuracy Detection for among ML models for PD Detection

TABLE V

ML Model	Percentage of Accuracy Contribution
XGBoost	13.3% – Highest performer
SVM-MFW	13.2%
EESDPD	12.5%
Extra Tree	12.5%
MASS-PCNN	12.3%
SVM	12.1%
KNN	12.1%
PLR	12.0% – Lowest performer

- Top Performers:
 - XGBoost (13.3%) and SVM-MFW (13.2%) are the leading models in terms of detection accuracy.
 - These models are often used in ensemble or feature-weighted approaches, which likely contributed to their higher performance.
- Moderate Performers:
 - EESDPD, Extra Tree, and MASS-PCNN range from 12.3% to 12.5%, showing balanced effectiveness.
- Lower Performers:
 - PLR (12.0%) had the lowest relative accuracy, but the difference across all models is marginal (within 1.3%).

IX. CHALLENGES AND ETHICAL CONSIDERATIONS

While AI enhances PD diagnosis, challenges remain:

- Limited labeled data and imbalanced datasets
- Interpretability of complex DL models
- Patient privacy and data protection

Addressing these will ensure safer and more trustworthy AI systems in healthcare.

X. CONCLUSION

AI provides revolutionary tools in the battle against Parkinson's Disease, facilitating early detection, enhanced monitoring, and superior patient outcomes. With ongoing research and ethical implementation, AI is set to become an essential component of neurological care. Recent progress in ML and DL has changed how PD is diagnosed by making it more accurate and objective. Research shows that machine learning (ML) and deep learning (DL) can find Parkinson's disease (PD) in a lot of different types of data, such as EEG signals, MRI images, and speech. Using neural networks like ANNs and CNNs, these methods have been better at telling the difference between PD patients and HC patients. For instance, they used EEG signals to accurately group PD patients and CNNs on MRI images to find structural changes that were linked to PD. Additionally, utilized speech data to identify voice biomarkers linked to PD. In this study, we performed a comprehensive assessment of the recent literature utilizing ML and DL techniques for PD analysis. Our synthesis encompassed a summary of the preprocessing techniques employed with the accessible datasets in these studies.

Contributions from the author H.R. wrote most of the book and did the research for it. During the writing of the manuscript, M.A. gave conceptual advice and kept an eye on things. The manuscript was read by all of the authors. Cash The Natural Sciences and Engineering Research Council of Canada (NSERC) gave this research money with the reference number RGPIN-2024-05287, which made it possible. Getting to data Not relevant. Availability of code Not relevant. Availability of materials Not relevant

REFERENCES

- [1] Europe P. Parkinson's Statistics. 2024. <https://parkinsonseurope.org/facts-and-figures/statistics/>. Accessed 2 Oct 2024.
- [2] Mayo Clinic. Parkinson's disease. 2024. <https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055>. Accessed 2 Oct 2024.
- [3] Organization WH. Parkinson disease. Online, 2024. <https://www.who.int/news-room/fact-sheets/detail/parkinson-disease>. Accessed 2 Oct 2024.
- [4] Tysnes O-B, Storstein A. Epidemiology of Parkinson's disease. *J Neural Transm.* 2017;124:901–5.
- [5] Dorsey ER, Elbaz A, Nichols E, Abbasi N, Abd-Allah F, Abdelalim A, Adsuar JC, Ansha MG, Brayne C, Choi J-YJ, et al. Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet Neurol.* 2018;17(11):939–53.
- [6] Tolosa E, Garrido A, Scholz SW, Poewe W. Challenges in the diagnosis of Parkinson's disease. *Lancet Neurol.* 2021;20(5):385–97.
- [7] Willis A, Roberts E, Beck J, Fiske B, Ross W, Savica R, Van Den Eeden S, Tanner C, Marras CM. Incidence of Parkinson disease in North America. *NPJ Parkinson's Dis.* 2022;8(1):170.
- [8] Parkinson's Foundation <https://www.parkinson.org/>. Accessed 14 June 2024.
- [9] Parkinson's Foundation: Understanding Parkinson's: Statistics. <https://www.parkinson.org/understanding-parkinsons/statistics>. Accessed 14 June 2024.
- [10] Little M, McSharry P, Hunter E, Spielman J, Ramig L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nature Precedings*, 2008;1–1
- [11] Dimauro G, Di Nicola V, Bevilacqua V, Caivano D, Girardi F. Assessment of speech intelligibility in Parkinson's disease using a speech-to-text system. *IEEE Access.* 2017;5:22199–208. <https://doi.org/10.1109/ACCESS.2017.2762475>.
- [12] Dimauro G, Girardi F. Italian Parkinson's voice and speech. *IEEE Dataport* 2019. <https://doi.org/10.21227/aw6b-tg17>
- [13] Kursun O, Sakar B, Isenkul M, Sakar C, Sertbas A, Gorgen F. Parkinson Speech Dataset with Multiple Types of Sound Recordings. *UCI Machine Learning Repository* 2014. <https://doi.org/10.24432/C5NC8M>
- [14] Tsanas A, Little M. Parkinsons Telemonitoring. *UCI Machine Learning Repository.* 2009. <https://doi.org/10.24432/C5ZS3N>
- [15] Sakar C, Serbes G, Gunduz A, Nizam H, Sakar B. Parkinson's Disease Classification. *UCI Machine Learning Repository* 2018. <https://doi.org/10.24432/C5MS4X>
- [16] Isenkul M, Sakar B, Kursun O, et al. Improved spiral test using digitized graphics tablet for monitoring parkinson's disease. In: *The 2nd International Conference on E-health and Telemedicine (ICEHTM-2014)*, 2014;5:171–175