Advanced Pathways in Electrical, Communication, and Automation: Reconfigurable Systems, Smart Energy, and AI for Industry 5.0 – Dr. A. Punitha et al. eds

© 2026 Taylor & Francis Group, London, ISBN 978-1-041-30297-1

4
Neurodegenerative Disorder Prognostics via Machine Learning: Predictive Modelling for Parkinson's Disease

Arunkumar S.[1]

Associate Professor, Department of Mechanical Engineering, Vels Institute of Science, Technology & Advanced Studies, Chennai, India

Malathi R.[2]

Assistant Professor, Department of Computer science and Engineering, C. K. College of Engineering and Technology, Cuddalore, India

Ranjith P.[3]

Assistant Professor, Department of Mechanical Engineering, Gojan School of Business and Technology, Chennai, India

Cyril Prasanna Ra[4]

Professor, Department of Electronics & Communication Engineering, Cambridge Institute of Technology, Bengaluru, India

Saranraj I.[5]

Associate Professor, Department of Mechanical Engineering, Mahendra College of Engineering, Salem, India

Kowsalya L.[6]

Assistant Professor, Department of CSBS, Gojan School of Business and Technology, Chennai, India

[1]gct.arunkumar@gmail.com⏎

[2]malathi022@gmail.com⏎

[3]ranjith.v@gsbt.edu.in⏎

[4]cyril@cambridge.edu.in⏎

[5]saran13dr@gmail.com⏎

[6]org.vipstech@ gmail.com ⏎

DOI: 10.1201/9781003773962-4

Abstract: In modern times, we commonly observe Parkinson's disease (PD) difficulties, which have a significant probability of contributing to fatalities. PD is also the reason for the elevated worldwide rate of death. Approximately six million folks around the globe are plagued with this illness. It transpires via the demise of dopaminergic-establishing synapses and typically impacts persons around the age of 60. Parkinson's illness cannot possess a treatment; however, early discovery may be likely to decrease the illness's course. The patients are suffering from vocal cord problems. Lack of speech is an early sign of PD. This research emphasizes the creation of Parkinson's condition identification systems employing machine learning (ML) methods that enable precise estimation and smart decisions and correct assumptions for analyzing information. According to the findings, our recommended approach trumps other strategies.

In anticipation of accomplishing our objective, we have begun using a cooperative teaching technique. The proposed model improves current Traditional model these include Decision Tree, Random Forest, Logistic Regression, Naïve Bayes, KNN, as well as CatBoost. We succeeded in getting an amazing testing accuracy that reached roughly 97.67% with CatBoost. The suggested algorithm is built on ML approaches to combine various models and datasets, giving actual time, exact diagnoses, as well as customized medical treatments. This study leverages organized datasets and powerful ML algorithms to give an approach for adaptable and efficient cardiac disease identification, possibly lowering fatality rates and boosting therapeutic outcomes.

Keywords: Parkinson’s disease, Machine learning, CatBoost, Prediction

1. Introduction

Parkinson’s sickness that emerges as a consequence of chronic decline of dopamine, which a neural transmitters. PD is lot frequent in the aging mammalian populace , producing variations in stride as well as position that might raise the potential of tripping as well as lead to disability issues [1]. It happens owing to the destruction of neurons, leading to a reduction of catecholamine levels in the brain. Insufficient amounts of catecholamine hinder transmission within neural junctions, producing inefficient cognitive processes [2].

PD’s distinguishing behavioral signs, including tremors, rigidity, bradykinesia, and posture unpredictability, are connected to the degeneration of dopamine-producing neurons. Along with those, PD shows non-motor signs like constipation, hyposmia, REM sleep behavior disorder (RBD), neuropsychiatric indicators, and autonomic disorders, which may develop in the initial phases of an illness, frequently years before motor signs lead to diagnostic procedures [3]. These approaches accurately detect illness development in its beforehand phases, even with the faint as well as milder characteristics accessible in MRI scans, as well as reduce the natural inaccuracy inherent in clinician-based handwritten assessments. However, although using ML/DL computations to accelerate sickness diagnosis is beneficial, the systems need a substantial quantity of input for learning to categorize effectively [4, 5].

Through the combination of Paralysis Agitans illness diagnosis, classification, as well as prognosis, the paper’s unique paradigm enables individualized disease classification as well as prediction. A complicated model employing classifiers from Decision Tree, Random Forest, Logistic Regression, Naïve Bayes, K-Nearest Neighbors, as well as CatBoost combined with the generation of specific datasets like Parkinson’s Disease, exhibits the feasibility of correct predictions. To summarize our efforts in this paper:

- We obtained Parkinson’s Disease datasets from the Kaggle website and grouped them into multiple classes in this research.
- We compared the efficacy of our suggested techniques.

- Comparing our suggested models to those that presently exist.

2. Literature Review

This paper indicates the effectiveness of Plentiful classifications utilizing scrub-oriented and packagingoriented feature selection approaches. To strengthen the Parkinson's disease prognosis. Parameters including Correctness, Exactness, Sensitivity, and Balanced Score were compared with the original dataset. The wrapperbased feature selection obtained 88.33% using KNN [6]. The machine learning is used more to analyze the EEG data; it helps to predict the Parkinson's illness. Most of the research focuses on the test to achieve over 90% accuracy, even with small datasets and unstructured clinical details. The choice of feature and model is developed for good results, while the EEG cleaning steps are less important, making the automated prediction models faster and easier to develop [7].

Wuwang et al. [8] being aware of Parkinson's Illness (PI) critical to comprehending why it occurs and finding therapies. This work presented a model developed using deep learning incorporating premotor characteristics, including REM insomnia and olfactory loss, reaching 96.45% accuracy. The approach beat 12 machine learning methods, suggesting high promise for early PD identification while utilizing a short dataset. Mohsin Raza et al. [9]suggested IoT and machine learning system aids Parkinson's patients by providing ongoing tracking of electrophysiological and environmental conditions with minimum latency and effective sensor connection. It can manage up to 250 detectors and forecasts PD development with great accuracy using speech transcripts. While successful, it lacks interaction with inertial sensors and motion data, providing opportunities for further investigation.

The prompt indicator of Parkinson's Illness is to limit the death mortality. This model performs the contemporary learning techniques, comprising a decision tree, SVM, KNN, random forest, as well as XGBoost, attaining 94.87% accuracy, 81.99% MCC, as well as 94.52% F1 score. It indicates that it has a larger accuracy performance in early prediction compared to other approaches [10]. Naresh Alapati et al. [11] conduct the comparison of accuracy, recall, as well as F1 score, revealing that the Random Forest (RF) technique beats KNN in predicting Parkinson's Disease. The random forest obtained 93% accuracy compared to KNN, which reached 89%

effectiveness with less training time. Thus the RF performs best and boosts Parkinson's disease prediction.

Big data analytics is changing medical care to make treatments more accurate, clear, and low-cost using better technology and large amounts of data. This research introduces a system in which multiple classifiers are relied on to anticipate the onset of Parkinson's and come across it promptly. The apparatus enhanced precision and effectiveness, leading to better medical results. This approach will bring big benefits in productivity, revenue, and efficiency for healthcare companies [12]. Srishti Grover et al. [13] constructed a DNN that can estimate degree of Parkinson's sickness. The recommended DNN model demonstrated greater effectiveness when assessed versus substitute methods. It also appears that the segmentation determined by motor UPDRS score prevails over the sorting relying on broadly UPDRS result, and subsequently, may be extrapolated to represent a better metric for susceptibility foresight. Despite we relied on a collection of 5875 instances, the credibility of this evaluation may be higher broadened by doing it on a wider set of data, bringing increased lots of circumstances of each degree tier together. Databases employed for patients' Data on speech alongside additional client information including stride and writing communication elements.

The research demonstrates that machine learning models and picking features approaches are crucial to increasing Parkinson's illness prediction success, with Random Forest, KNN[14], and deep learning particularly especially successful. Combining EEG data and IoT technologies for ongoing monitoring has also shown potential for prompt identification. These advancements assist in developing speedier and more exact diagnoses, leading to better healthcare outcomes.

3. Proposed Methodology

A detailed description of the suggested research framework's architecture is offered as a process in "Fig. 4.1". The image below demonstrates a thorough knowledge of the structure and features of the proposed framework. The associated offers a full overview of the different components as well as structure of the proposed framework.

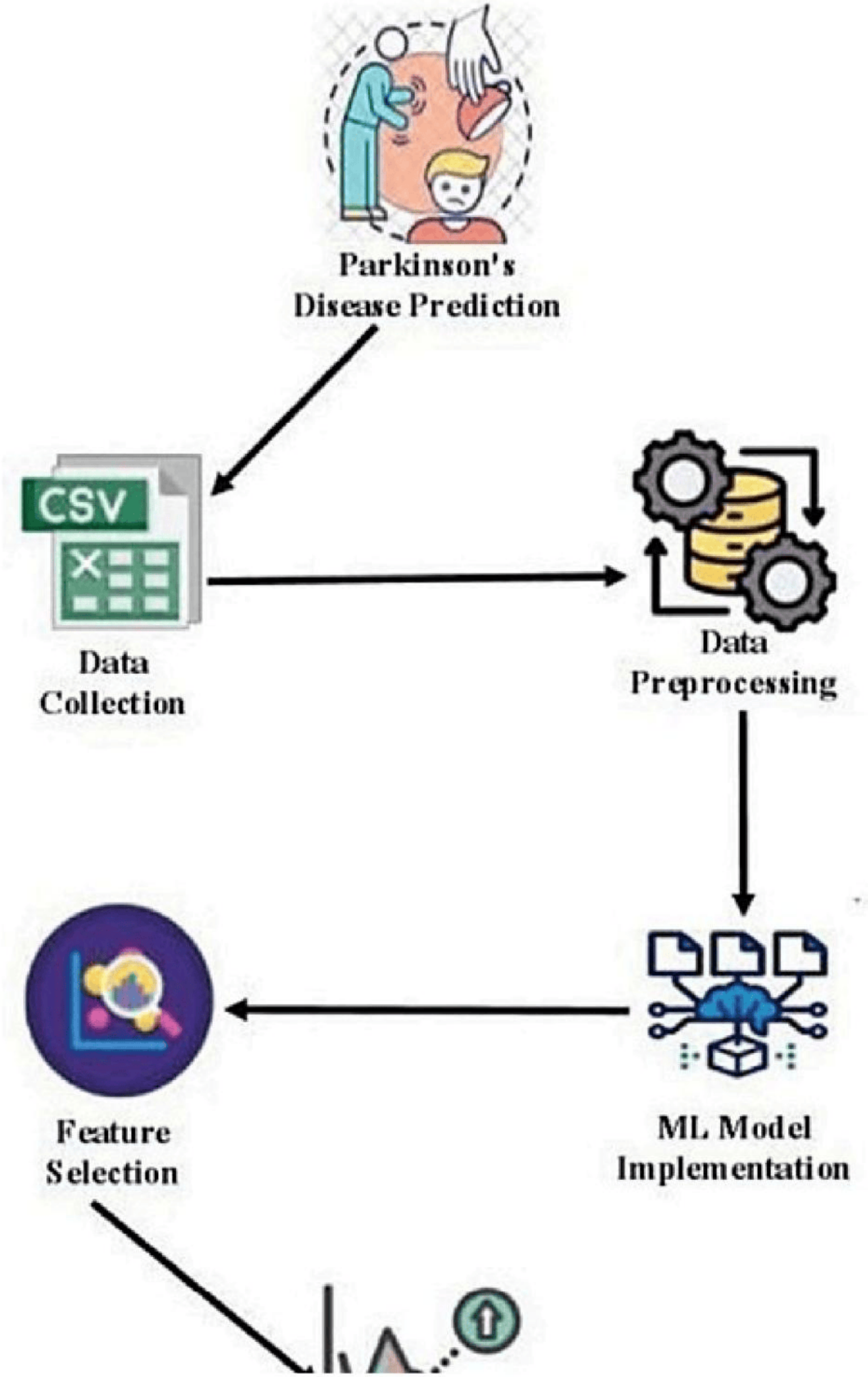
Parkinson's
Disease Prediction
CSV
Data
Collection
Data
Preprocessing
ML Model
Implementation
Feature
Selection

Fig. 4.1 Parkinson's prediction workflow diagram⏎

3.1. Parkinson's Illness Dataset Compilation

The dataset[15] was acquired from trusted online sources, medical facilities, and clinics. Key features like MDVP: Fo(Hz), MDVP: Fhi(Hz), MDVP: Flo(Hz), MDVP: Jitter(%), MDVP: Jitter(Abs), MDVP: Shimmer, MDVP: Shimmer(dB), NHR, and HNR are utilized toexamine voice qualities for Parkinson's illness diagnosis. "Figure 4.2" comprises both affected and non-affected people, with labels determining the illness state (1 for Not Affected, 0 for affected). The information collection was permitted by the ethical evaluation Board of each institution involved.

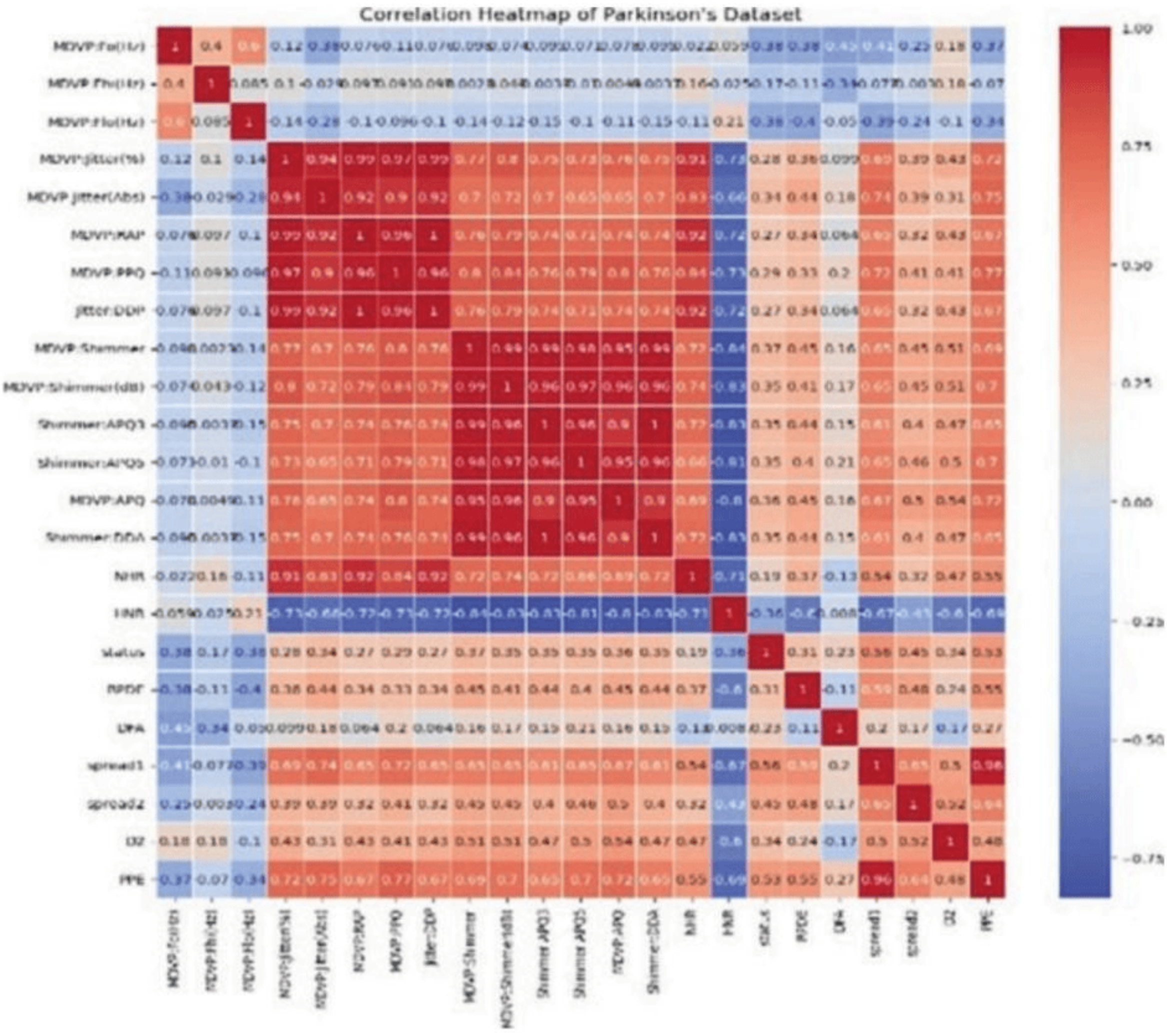

Fig. 4.2 Heatmap for parkinson's dataset

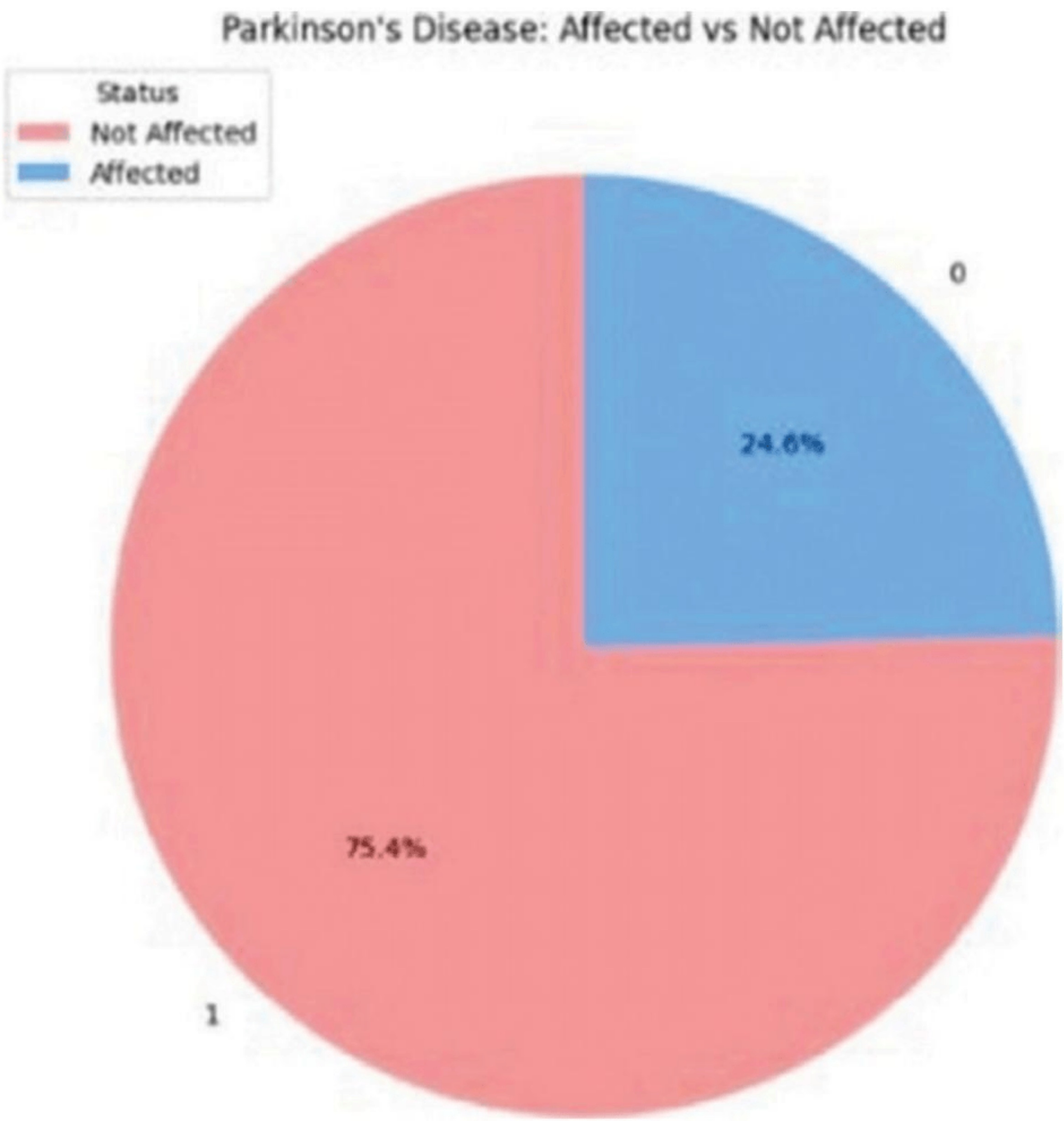

Fig. 4.3 Classification between overall parkinson's affected vs. non-affected

"Figure 4.4" explains the fundamental frequency analysis of MDVP: Fo(Hz) - This corresponds to the the core severity employed by the voice, representing the majority of the pitch generated during phonation. It indicates the fundamental frequency of vocal cord oscillation, MDVP: Fhi(Hz) - It refers to the peak fundamental frequency recorded throughout phonation, signifying the most extreme pitch attainable by the vocal cords, MDVP: Flo(HZ) - It represents the most modest fundamental frequency measured during phonation, revealing the least pitch the voice cords generate.

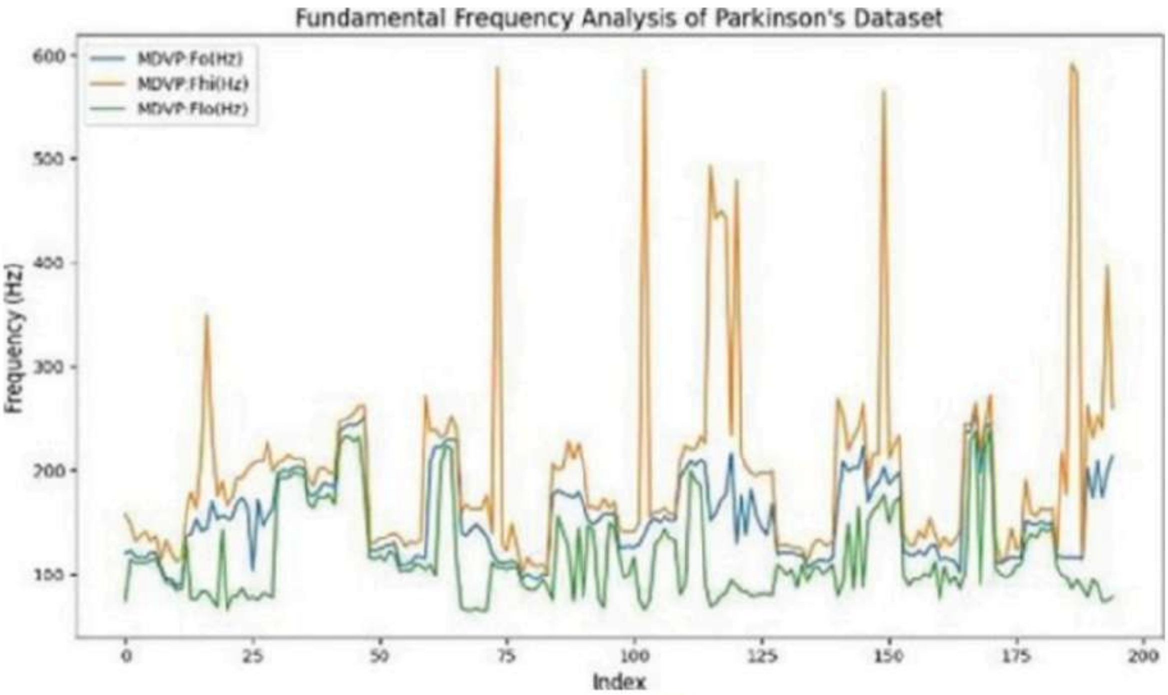

Fig. 4.4 Fundamental frequency analysis⏎

3.2. Model Selection

A variety of ML algorithms were used in our experimental study, including RF [14], DT [10], LR [15], NB [6], KNN [10], and CB.

Decision Tree (DT): Decision tree evaluation, a comprehensive forecasting technique, has multiple uses. Decision trees are frequently generated through a methodical approach that identifies methods to segment the data source based on numerous factors. It has emerged as one of the most extensively utilized and reliable strategies for supervised learning. The supervised learning approach known as decision trees is used for tasks involving regression and classification. The intent is to generate obvious decision guidelines utilizing information details while building a hypothesis that anticipates the eventual value of the sought variable. The decision tree is used for the "equation (1)" where the decision tree separates input at every node depending on a characteristic yj yi and limit *t* using the rule. This iterative algorithm separates the feature field into areas, with predictions made at the branching nodes.

$$f(y) = a \text{if} y_i \leq t \quad \{b \text{if} y_i > t\} \tag{1}$$

Random Forest (RF): RF is a classifier made up of several decision trees, which determines the type by selecting a combination of the outputs from every one of the trees. The "equation (2)" for generating a RF was created by Leo Breiman as well as Adele Cutler, as well as "Random

Forests" is their registered brand. The technique blends Breiman's "bagging" notion with the random selection of characteristics, implemented separately by Ho as well as Amit as well as Geman in order to produce a set of decision trees with regulated variance.

$$y^{\wedge} = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \tag{2}$$

Logistic Regression (LR): LR is a forecasting model when the class identifier or the objective is categorical. This variable possesses two categories, yes/no. For this Parkinson's Disease dataset, the classification will be whether the individual is afflicted by Parkinson's Disease or the individual is a non-Parkinson Disease. The "equation (3)" be used exclusively with two kinds of objective variables. ( )

$$PC_y \mid z| = \frac{PC_y * PzC_y}{P(z)} \tag{3}$$

Naïve Bayes (NB): NB is a Likelihood-based predictor that relies entirely Founded on Bayesian theory. It qualifies as Naïve owing to the hypothesis conveyed by the classifier relies on a powerful features autonomy presumption. In literary works, there are various varieties of NB: basic Naïve Bayes, Gaussian Naïve Bayes, Multinomial Naïve Bayes, Bernoulli Naïve Bayes, as well as Multidimensional Poisson Naïve Bayes , wherein the fundamental distinction amongst species is the method the value of the targeted class is estimated.

$$q(x = 1 \mid y \mid) = \frac{1}{1 + e^{-z}} \tag{4}$$

K-Nearest Neighbor (KNN): Although KNN relies on calculating boundaries, it is extremely dependent on the amount of data from training. This tactic would not return findings depending on assumptions about the learning set of information. The KNN approach is recognized as being fairly excellent at resolving handling unpredictable observations, despite this might necessarily be the overall scenario when you investigate alternative supervised learning methods.

$$y^{\wedge} = argmax_c \sum_{i=1}^{k} \Pi v_t = c \tag{5}$$

CatBoost: CatBoost (categorical Boosting) is a gradientboosting method intended for processing category information effectively. It employs hierarchical boosting to prevent excessive fitting and applies balanced trees for quicker training and improved adaptation. CatBoost efficiently preprocesses category data by turning it into mathematical models using targeted statistics. It enables GPU acceleration, making it particularly efficient for huge datasets.

$$F(x) = F_0(x) + \sum_{m=1}^{M} \gamma_m h_m \tag{6}$$

4. Result and Discussion

This work's findings of the study were acquired utilizing a PD dataset. The project aims to analyze the effectiveness of multiple ML algorithms in classifying as well as forecasting results within this area. A combination of models containing many ML techniques was built, and the accuracy rates of these methods are given in "Fig. 4.5". The ML algorithms investigated for the research were DT, RF, LR, Naive Bayes, KNN, and CatBoost. As the dataset size became bigger, substantial discrepancies in accuracy arose between the algorithms. "Table 4.1" presents a complete study of the accuracy rates reached by each approach, clearly indicating the higher performance of CatBoost. This work underlines the possibilities for hybrid optimization techniques in boosting the stability and precision of machine learning strategies in crucial fields, especially PD diagnosis. Our major contribution to the effort is establishing Actual Parkinson's condition datasets, specifically. The PD dataset, developed for identifying patterns in twenty-four PD categories, demonstrated great accuracy in Catboost. The PD dataset, including unaffected people, permits binary categorization. The PD dataset, including affected and unaffected cases, generated the largest performance in CatBoost at 97.67%

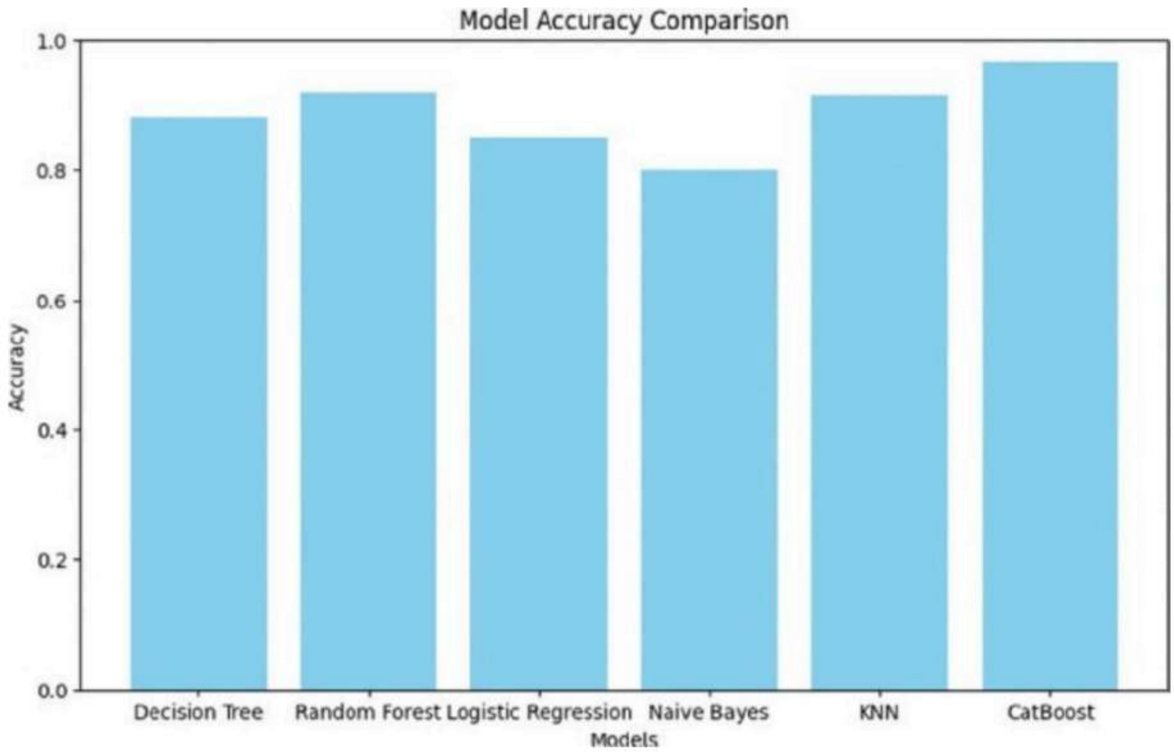

Fig. 4.5 Algorithm comparison for parkinson's disease prediction⏎

Table 4.1 Accuracy differentiation report⏎

Techniques	Accuracy (%)
Decision Tree	92%
Random Forest	93%
Logistic Regression Naïve Bayes	83% 76%
KNN	91%
CatBoost	**97%**

4.1. Performance Evaluation using Confusion Matrix

A Confusion matrix is an expression of data that is frequently used to classify issues. That data chart exhibits scenarios of genuine lessons as columns as well as instances of predicted courses as column. "Fig. 4.6" represent the table features 4 amounts: the quantity of Outcomes of Accurate Positive (AP), Accurate Negative (AN), Erroneous Positive (EP), and Erroneous Negative (EN). What proportion of encounters are truly associated with the intended class as well as have been accurately identified as such (Correct positives). The proportion of information items that were properly categorized as not matching the target class is called the Correct negative (CN) rate. This would comprise metrics that would be utilized to evaluate the model's efficacy, covering precision, recall, and accuracy, and even the

F1 score. When assessing how well binary classification models work, accuracy is a frequently used parameter. Based on each of the predictions the model produces, it estimates the number of accurate projections. You can use the following "equation (7)" to determine accuracy.

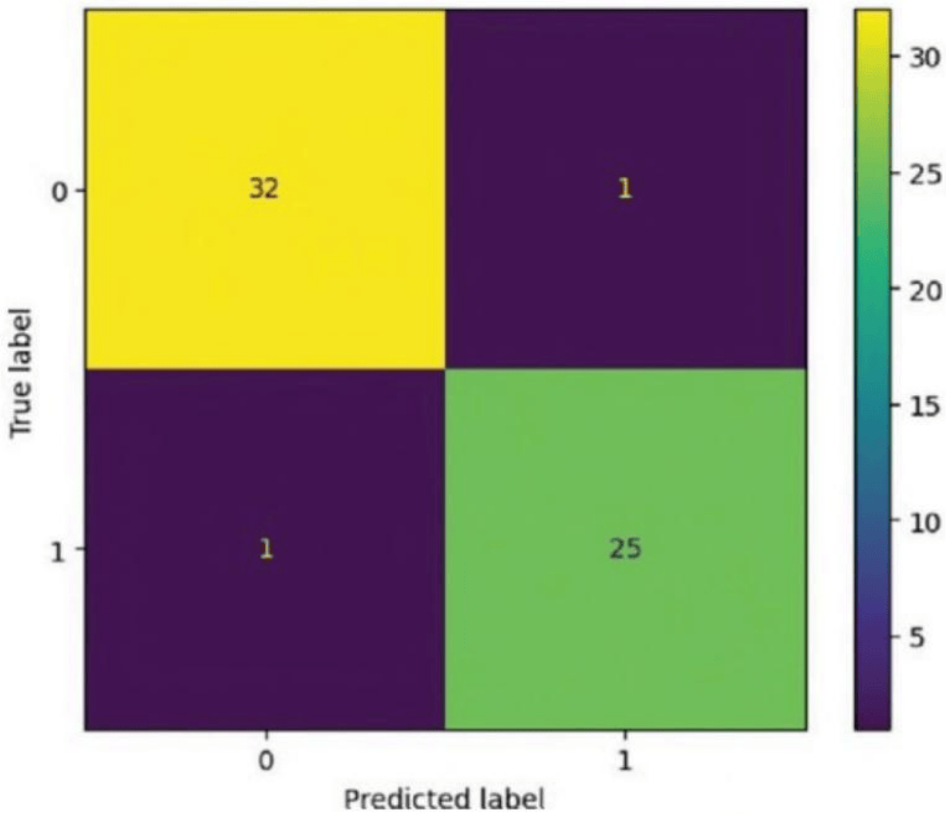

Fig. 4.6 Confusion matrix for CatBoost algorithm⏎

$$AF(x) = F_0(x) + \sum_{m=1}^{M} \gamma_m \hbar_m \tag{7}$$

5. Conclusion

The severity of PD can be recognized by utilizing a CatBoost model of ML approaches. As proven by our experimental outcomes, the CatBoost technique has obtained greater accuracy (97%) in contrast to ML classifiers such as LR (83%), Naïve Bayes (76%), KNN (91%), RF (93%), Dt (92%), and others. As the dataset develops, the algorithm has done an adequate job of assessing the learned material and properly categorizing it. In this work, five algorithms are tailored, trained, and evaluated using five different models, including DT, RF, LR, Naive Bayes, KNN, and CatBoost, both with and without hyperparameter tweaking strategies. Their accuracies are compared with the current approaches. Significantly modifying the specifications of CatBoost with testing accuracy attained 97%. CatBoost offers the maximum testing accuracy of these five techniques, at 97%. The CatBoost classifier is the best hyperparameter for accuracy. In the future, the study will concentrate on applying distinct machine learning advancements and growing pattern-picking strategies to enhance the accuracy and effectiveness of PD prediction. Optimization techniques will be incorporated to increase the performance of prediction systems, while efforts will also extend to post-diagnostic elements, including regulating and treating Parkinson's illnesses.

References

1. Radha N, Sachin Madhavan RM, Sameera holy S, "Parkinson's Disease Detection using Machine Learning Techniques," *International Journal of Early Childhood Special Education (INT-JECSE),* vol. 30, no. 2, p. 543, Jan. 2021, doi: 10.24205/03276716.2020.4055.⏎
2. A. Govindu and S. Palwe, "Early detection of Parkinson's disease using machine learning," *Procedia Computer Science,* vol. 218, pp. 249-261, Jan. 2023, doi: 10.1016/j.procs.2023.01.007.⏎
3. T. Bhattacharya, K. T. Thomas, and L. Mathew, "*Parkinson's Disease Progression Prediction using Advanced Machine Learning Techniques,*" ICEECT, pp. 1-5, Aug. 2024, doi: 10.1109/iceect61758.2024.10739044.⏎
4. S. Priyadharshini et al., "A Comprehensive framework for Parkinson's disease diagnosis using explainable artificial intelligence empowered machine learning techniques," *Alexandria Engineering Journal,* vol. 107, pp. 568-582, Jul. 2024, doi: 10.1016/j.aej.2024.07.106.⏎
5. A. Schrag, "How valid is the clinical diagnosis of Parkinson's disease in the community?," *Journal of Neurology Neurosurgery & Psychiatry,* vol. 73, no. 5, pp. 529-534, Oct. 2002, doi: 10.1136/jnnp.73.5.529.⏎
6. F. Saeed et al., "Enhancing Parkinson's disease prediction using machine learning and feature selection methods," *Computers, Materials & Continua/Computers, Materials & Continua (Print),* vol. 71, no. 3, pp. 5639-5658, Jan. 2022, doi: 10.32604/cmc.2022.023124.⏎

7. A. M. Maitín, A. J. García-Tejedor, and J. P. R. Muñoz, "Machine Learning Approaches for Detecting Parkinson's Disease from EEG Analysis: A Systematic Review," *Applied Sciences,* vol. 10, no. 23, p. 8662, Dec. 2020, doi: 10.3390/app10238662.⏎
8. W. Wang, J. Lee, F. Harrou, and Y. Sun, "Early detection of Parkinson's disease using deep learning and machine learning," *IEEE Access,* vol. 8, pp. 147635-147646, Jan. 2020, doi: 10.1109/access.2020.3016062.⏎
9. M. Raza, M. Awais, N. Singh, M. Imran, and S. Hussain, "Intelligent IoT framework for indoor healthcare monitoring of Parkinson's disease patient," *IEEE Journal on Selected Areas in Communications,* vol. 39, no. 2, pp. 593-602, Sep. 2020, doi: 10.1109/jsac.2020.3021571.⏎
10. "Early Warning Signs of Parkinson's Disease Prediction Using Machine Learning Technique", *Journal of Pharmaceutical Negative Results*, pp. 4784-4792, Dec. 2022, doi: 10.47750/pnr.2022.13.S10.579.
11. N. Alapati, N. Anusha, P. Joharika, N. J. Jerusha, and P. Tanuja, "Prediction of Parkinson's Disease using Machine Learning," *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*, pp. 1357-1361, Mar. 2023, doi: 10.1109/icears56392.2023.10085443.⏎
12. Shamli, N., and B. Sathiyabhama. "Parkinson's Brain disease prediction using big data analytics." *International Journal of Information Technology and Computer Science (IJITCS)* 8.6 (2016): 73.⏎
13. S. Grover, S. Bhartia, N. Akshama, A. Yadav, and S. KR, "Predicting severity of Parkinson's disease using deep learning," *Procedia Computer Science,* vol. 132, pp. 1788-1794, Jan. 2018, doi: 10.1016/j.procs.2018.05.154.⏎
14. A. Deepak kumar, N. Revathi, S. IrinSherly, R. Lalitha, R. Vinston Raja., *Innovative Time Series-Based Ecg Feature Extraction For Heart Disease Risk Assessment Journal of Theoretical and Applied Information Technology*, 15th November 2023 -- Vol. 101. No. 21-- 2023⏎
15. Selvan C, Jothi Prabha Appadurai, Habib Kraiem and Rathish C R, (2025), Machine Learning-Enabled Optimization of a Graphene Coated Terahertz Metasurface for Isoquercitrin Biosensing, *Journal of The Electrochemical Society*, Volume 172, Number 6, https://iopscience.iop.org/article/10.1149/1945-7111/ade347/meta⏎

16. Kalaimani, G., Kavitha, G., Chinnaiyan, S. et al. Optimally configured generative adversarial networks to distinguish real and AI-generated human faces. *SIViP* 18, 7921-7938 (2024). https://doi.org/10.1007/s11760-024-03440-6
17. Ghattamaneni, Dileep Kumar. "Leveraging Artificial Intelligence and Automation for Effective Lot Disposition and Quality Control in Semiconductor Manufacturing." *Excel International Journal of Technology, Engineering & Management*, vol. 6, no. 4, 2019, pp. 29-35. https://exceljournals.org.in/detail.php?id=758
18. Vinston Raja R, Deepak Kumar A, PrabuSankar N, Chidambarathanu K, Thamarai I, Krishnaraj M, IrinSherly S., Comparative Evaluation of Cardiovascular Disease Using MLR and RF Algorithm With Semantic Equivalence., *Journal of Theoretical and Applied Information Technology*., 30th September 2023 -- Vol. 101. No. 18-- 2023
19. PrabuSankar, N. Jayaram, R. IrinSherly, S. Gnanaprakasam, C. Vinston Raja, R. Study of ECG Analysis based Cardiac Disease Prediction using Deep Learning Techniques, *International Journal of Intelligent Systems and Applications in Engineering*, 2023, 11(4), pp. 431-438
20. Vinston Raja, R. and Ashok Kumar, K. 'Financial Derivative Features Based Integrated Potential Fishing Zone (IPFZ) Future Forecast'. *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 3, pp. 3637-3649, 2023.

Note: All the figures and the table in this chapter were made by the authors.