ISSN: 2977-0033

Research Article

Journal of Clinical Research and Case Studies

Comparative Analysis of Machine Learning Algorithms for Accurate Parkinson's Disease Detection

Ravi Shankar Sah[1], Saugat Kafle[1], Sudip Raj Khadka[1] and Smriti KC[2*]

[1]*Samriddhi College, Lokanthali, Bhaktapur, Nepal*
[2]*Padma Kanya Campus, Bagbazaar, Kathmandu, Nepal*

[*]Corresponding author
Smriti KC, Padma Kanya Campus, Bagbazaar, Kathmandu, Nepal.

Received: January 08, 2025; **Accepted:** January 16, 2025; **Published:** January 21, 2025

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder affecting millions globally. Early diagnosis is vital in managing the condition and improving patient outcomes. This study investigates the use of machine learning algorithms to detect Parkinson's disease based on various biomedical features. A comparative analysis was performed using several classification algorithms, including AdaBoostClassifier, GradientBoostingClassifier, KNeighborClassifier, LGBMClassifier, and Random Forest Classifier, to identify patterns in predictive outcomes. Among these, the KNeighborClassifier achieved the highest accuracy of 95%. The study further compares precision, recall, and F1 scores across the algorithms, highlighting the potential of machine learning in clinical applications for Parkinson's disease detection. Additionally, the research emphasizes the importance of exploring advanced deep-learning techniques to enhance prediction accuracy.

Keywords: Parkinson's Disease, Machine Learning, Classification Algorithms, Predictive Modeling

Introduction

Parkinson's disease (PD) is a chronic and progressive neurodegenerative disorder that profoundly affects millions of individuals worldwide [1]. It is marked by a combination of motor and non-motor symptoms, which make its diagnosis and management particularly challenging. The motor symptoms include hallmark features such as resting tremors, muscle rigidity, bradykinesia (slowness of movement), and postural instability, all of which significantly impair mobility and quality of life. On the other hand, the non-motor symptoms of PD, which often precede motor manifestations, include cognitive impairments like memory loss and executive dysfunction, which can escalate to dementia in the later stages. Additionally, the disease is frequently associated with sleep disturbances, autonomic dysfunction, and mood disorders such as anxiety and depression, further contributing to its debilitating nature.

One of the greatest challenges in managing Parkinson's disease lies in its diagnosis. Currently, there is no definitive diagnostic test for PD, and the assessment relies heavily on clinical observation and patient-reported symptoms. These methods are inherently subjective and are often unable to detect the disease at its earliest stage when interventions could be most effective. By the time motor symptoms become apparent, significant neuronal damage particularly in the dopaminergic neurons of the substantia nigra has often already occurred. This diagnostic delay complicates treatment and diminishes the opportunity to slow disease progression or improve patient outcomes. Given these limitations, there is an urgent need for objective, reliable, and early diagnostic tools that can aid healthcare professionals in identifying PD before irreversible damage sets in.

In recent years, machine learning (ML) has emerged as a transformative technology with the potential to address these diagnostic challenges [2,3]. ML, a subset of artificial intelligence, uses advanced algorithms to analyze large and complex biomedical datasets, uncovering subtle patterns and

Citation: Ravi Shankar Sah, Saugat Kafle, Sudip Raj Khadka, Smriti KC. Comparative Analysis of Machine Learning Algorithms for Accurate Parkinson's Disease Detection. J Clin Res Case Stud. 2025. 3(1): 1-9. DOI: doi.org/10.61440/JCRCS.2025.v3.60

relationships that traditional statistical approaches may overlook. Unlike conventional diagnostic methods, ML models are capable of integrating data from diverse sources, such as speech recordings, motor performance assessments, wearable sensor data, and neuroimaging results, to make more accurate and earlier predictions about the presence of Parkinson's disease. Research in this domain has demonstrated promising results, with various ML techniques outperforming traditional approaches in terms of sensitivity and specificity [4,5]. By leveraging these capabilities, ML holds the potential to revolutionize the diagnostic landscape of PD, enabling clinicians to detect the disease earlier, monitor its progression more effectively, and tailor interventions to individual patient needs [6].

The contributions of this study are summarized in what follows:

1. To assess and compare the predictive accuracy of a range of machine learning models for detecting Parkinson's disease.
2. To conduct a comprehensive performance analysis of these models using metrics such as accuracy, sensitivity, specificity, and F1-score.
3. To explore the black-box approach of different ML models and find the major contributing features

Related Works

Sriram et al. implemented Bayes Net, Naive Bayes, Logistic, Simple Logistic, KStar, ADTree, J48, LMT, and Random Forest, achieving accuracy of 80.00%, 69.23%, 83.66%, 84.61%, 89.74%, 86.15%, 80.51%, 86.15%, and 90.26%, respectively [7]. The parallel coordinates showed higher variation in Parkinson's disease, highlighted in red. Various attributes, such as Fo, Flo, Jitter (%), Jitter (Abs), Jitter, Shimmer, Shimmer(dB), Shimmer: APQ3, Shimmer: APQ5, APQ, and Shimmer: DDA, were analyzed using sieve graphs showing inter and intra connections with healthy and diseased data. The ROC plots for Majority, k-nearest neighbor (k-NN), and SVM classification algorithms indicated that k-NN achieved an accuracy of 82.5%, while SVM achieved 88.9%. Tiwari et al. implemented Logistic Regression, Decision Tree, SVM, KNN, XGBoost, and Bagging algorithms, with test accuracies of 79%, 90%, 92%, 95%, 95%, and 92%, respectively, on 195 samples from 31 people, 23 with Parkinson's disease and 8 from the control group [8]. XGBoost achieved the highest accuracy rate of 95% and a training accuracy rate of 100%.

Challa et al. implemented Multilayer Perceptron, BayesNet, Random Forest, Boosted Logistic Regression, Boosted Trees, SVM, and Logistic Regression, with testing accuracies of 95.45%, 96.02%, 96.59%, 97.17%, 95.08%, 93.12%, 96.67%, and 97.14%, respectively. Boosted Logistic Regression gave the best performance with a 97.16% accuracy and 98.9% area under the ROC curve (AUC) [9].

Yadav et al. tested the parkinson disease Relying on chi2 test, extra trees classifier, and correlation matrix as feature extraction strategies and relying on Decision Trees, K-Nearest Neighbors, Random Forests, Bagging, AdaBoosting, and Gradient Boosting as supervised AI calculations for permutation calculations [10]. The calculation is based on obtaining higher classifier accuracy, as well as ROC curve accuracy. Three conspicuous component selection strategies allow each of the 23 features to select 10 best-performing features. The DT classifier has a higher accuracy of 94.87% in a dataset with 23 attributions, just like a dataset with 11 features. These results are also checked by ROC curve (AUC=98.7%). This calculation significantly separates PD patients from patients at the individual level, thus ensuring the use of computer-based findings in clinical practice.

Mittal et al. proposed a novel approach using data partitioning with feature selection algorithm principal component analysis (PCA) for Parkinson's disease classification [11]. In the proposed approach, the dataset has been divided into three equal parts, and two classes (healthy and Parkinson's disease) have been validated for individual data with different classifiers based on acoustic features. To improve the performance of classifying algorithms Principal Component Analysis (PCA) has been used. The minority and majority classes were obtained by applying the data set partition approach to the dataset of healthy and Parkinson's disease subjects. The three equal partitions were composed of healthy and PD classes. PCA was used for feature selection. We used three different classifiers to classify all data partitions, including the weighted k-NN (nearest neighbor, wkNN), Logistic Regression (LR), and Medium Gaussian Kernel Support Vector Machine (MGSVM). The classification accuracy of 74.2%, 85.0% and 82.1% was achieved using the Logistic algorithm, SVM with Gaussian, and weighted k-NN classifiers. The combination of classifiers, data partition and feature selection (first case) achieved a classification accuracy of 80%,87.63% and 89.23% respectively. In the second case, 85.2%, 89.36% and 90.3% accuracy with data partition and feature selection are obtained respectively. The results show that the proposed methodology could be used for Parkinson's disease classification.

johri et. al impLIMEnted two neural network-based models namely, VGFR Spectrogram Detector and Voice Impairment Classifier have been introduced, which aim to help doctors and people in diagnosing disease at an early stage [12]. An extensive empirical evaluation of CNNs has been implemented on largescale image classification of gait signals converted to spectrogram images and deep dense ANNs on the voice recordings to predict the disease. The experimental results indicate that the proposed models outperformed the existing state of the arts regarding accuracy. The classification accuracy on the VGFR spectrogram detector is recorded as 88.1% while the voice impairment classifier has shown 89.15% accuracy.

Wang et al. impLIMEnted an innovative deep-learning technique to early uncover whether an individual is affected with Parkinson Disease or not based on premotor features [13]. Specifically, to uncover PD at an early stage, several indicators have been considered in this study, including rapid eye movement and olfactory loss, Cerebrospinal fluid data, and dopaminergic imaging markers. A comparison between the proposed deep learning model and twelve machine learning and ensemble learning methods based on relatively small data including 183 healthy individuals and 401 early PD patients shows the superior detection performance of the designed model, which achieves the highest accuracy, 96.45% on average. Besides detecting the PD, they also provided the feature importance on the PD detection process based on the Boosting method.

Methodology

The workflow begins by loading and preprocessing the data, which includes encoding the target variable ('status'), splitting the data into training and testing sets, and scaling features using StandardScaler. This ensures data is in a suitable format for training machine learning models. A correlation heatmap is generated to visualize the relationships between features. Five different machine learning classifiers include gradient boosting, AdaBoost classifier, K-Nearest Neighbor (KNN) classifier, LightGBM classifier and Random Forest (RF). Each model's performance is assessed using different performance evaluation metrics. Local Interpretable Modelagnostic Explanation (LIME) is also used to explain individual predictions and understand the important features [14].

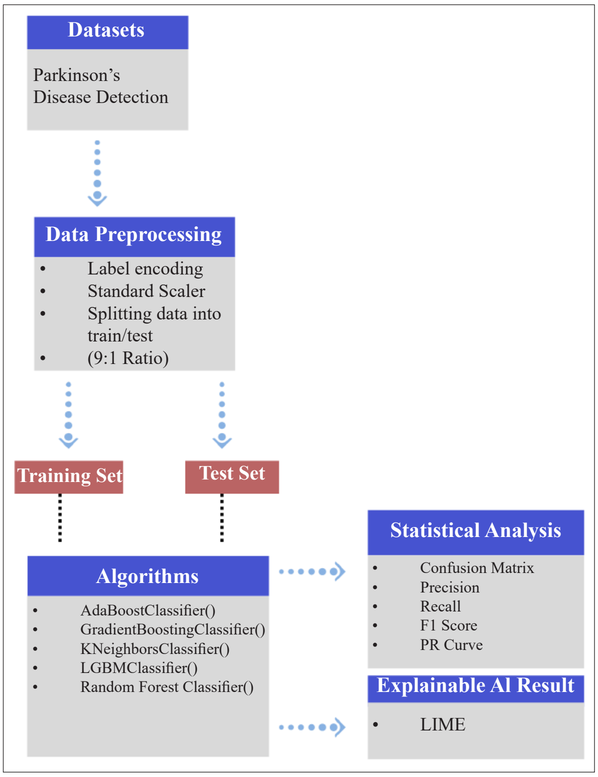


Figure 1: Methodology

Dataset Details

The Parkinson's Disease Voice Measurements dataset, as introduced by Little et al, comprises a comprehensive collection of biomedical voice features derived from individuals both with and without Parkinson's disease [15]. Each entry in the dataset corresponds to a voice recording and is characterized by a range of features, including fundamental frequency (Fo), jitter, shimmer, noise-to-harmonic ratio (NHR and HNR), and various nonlinear measures. The target variable, labeled as" status," serves as a binary indicator denoting the presence (1) or absence (0) of Parkinson's disease. For analysis, the dataset undergoes preprocessing steps that include the removal of the non-relevant 'name' column, verification and handling of missing values, and label encoding of the 'status' column. The resulting cleaned dataset is subsequently utilized to train machine learning models to classify and predict Parkinson's disease based on vocal attributes. A detailed description of the dataset is provided in Table 1.

Feature Important Analysis

The important features are typically derived from a predictive model trained on the dataset. These features can highlight patterns and relationships between features and the target, enabling data-driven decisions for feature selection and model optimization [16]. In Figure 2, the correlations between different features in the dataset are visualized, illustrating their interdependencies and contributions to the predictive task.

Table 1: Description of the dataset

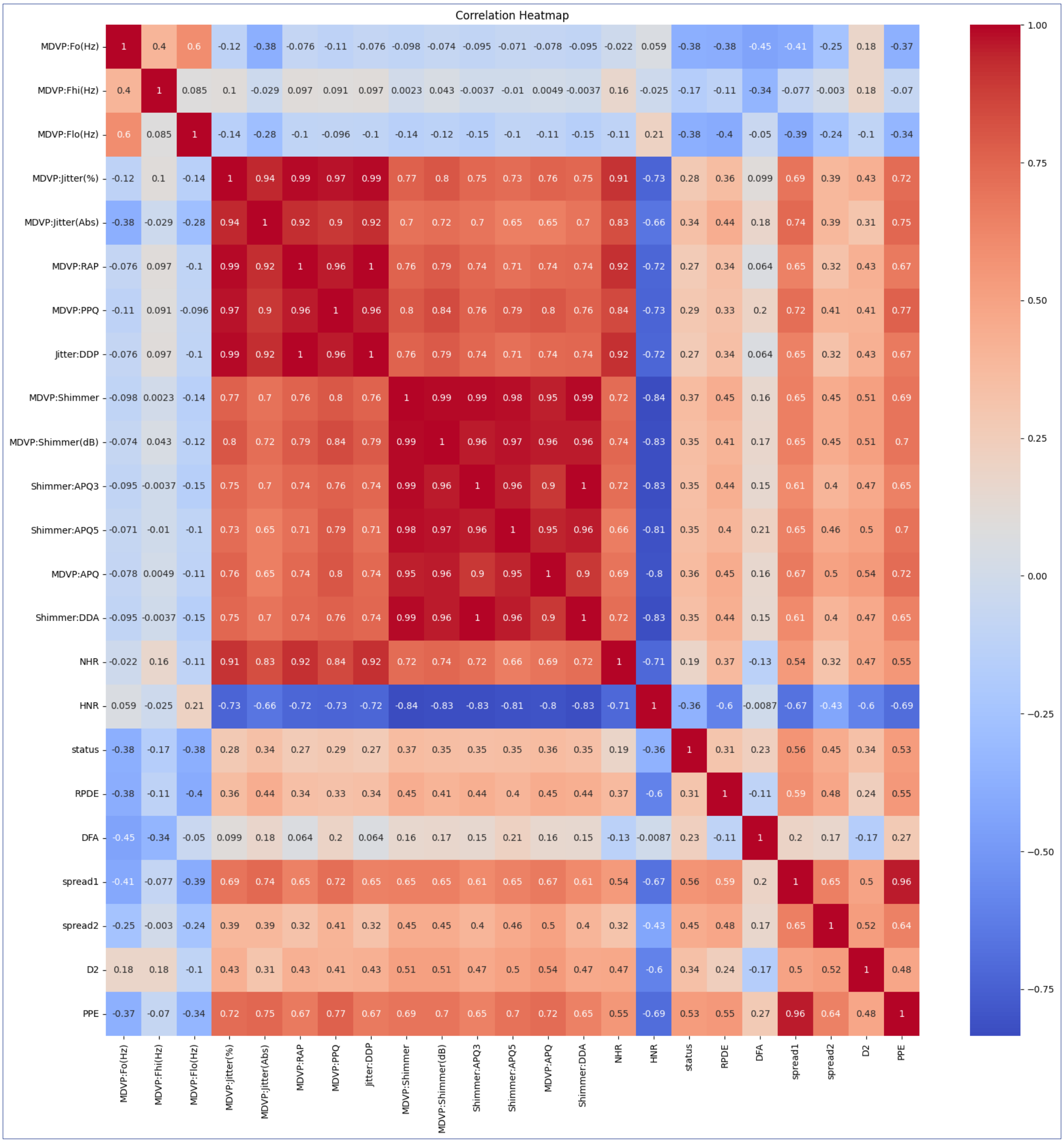
Attribute	Description
MDVP:Fo(Hz)	Average vocal fundamental frequency
MDVP:Fhi(Hz)	Maximum vocal fundamental frequency
MDVP:Flo(Hz)	Minimum vocal fundamental frequency
MDVP: Jitter(%), MDVP: Jitter(Abs), MDVP: RAP, MDVP: PPQ, Jitter: DDP	Several measures of variation in fundamental frequency
MDVP: Shimmer, MDVP: Shimmer(dB), Shimmer: APQ3, Shimmer: APQ5, MDVP: APQ, Shimmer: DDA	Several measures of variation in amplitude
PDE, D2	Two nonlinear dynamical complexity measures
NHR, HNR	Two measures of the ratio of noise to total components in the voice
DFA	Signal fractal scaling exponent
Spread1, Spread2, PPE	Three nonlinear measures of fundamental frequency variation
Status	Health status of the subject: (1) Parkinson's, (0) healthy

LIME

LIME operates by selecting a target instance x and model f, then generating perturbed instances around x. The predictions f(x) and are obtained, and interpretable features zi are derived from these perturbed instances. An interpretable model g(z) is then trained on pairs to approximate the behavior of f in the local vicinity of x. Analyzing g(z) involves examining the coefficients βi, which indicate the impact of features zi on predictions. The higher absolute value (βi) suggests a stronger influence of the respective features [17-19].

Evaluation Metrics and Experimental Setup

1. The performance of all implemented algorithms is measured using the following metrics: confusion matrix, precision, recall, F1-score, and PR curve.
2. **System Environment:** The machine learning models and XAI algorithms were developed in Python. Experiments were conducted on Google Colab, using an NVIDIA K80 GPU and 12 GB of RAM provided by Google. The Google Colab environment was configured with Python version 3.7, Keras version 2.5.0, and TensorFlow version 2.5.0.



Correlation Heatmap

	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	MDVP:Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	MDVP:APQ	Shimmer:DDA	NHR	HNR	status	RPDE	DFA	spread1	spread2	D2	PPE
MDVP:Fo(Hz)	1	0.4	0.6	-0.12	-0.38	-0.076	-0.11	-0.076	-0.098	-0.074	-0.095	-0.071	-0.078	-0.095	-0.022	0.059	-0.38	-0.38	-0.45	-0.41	-0.25	0.18	-0.37
MDVP:Fhi(Hz)	0.4	1	0.085	0.1	-0.029	0.097	0.091	0.097	0.0023	0.043	-0.0037	-0.01	0.0049	-0.0037	0.16	-0.025	-0.17	-0.11	-0.34	-0.077	-0.003	0.18	-0.07
MDVP:Flo(Hz)	0.6	0.085	1	-0.14	-0.28	-0.1	-0.096	-0.1	-0.14	-0.12	-0.15	-0.1	-0.11	-0.15	-0.11	0.21	-0.38	-0.4	-0.05	-0.39	-0.24	-0.1	-0.34
MDVP:Jitter(%)	-0.12	0.1	-0.14	1	0.94	0.99	0.97	0.99	0.77	0.8	0.75	0.73	0.76	0.75	0.91	-0.73	0.28	0.36	0.099	0.69	0.39	0.43	0.72
MDVP:Jitter(Abs)	-0.38	-0.029	-0.28	0.94	1	0.92	0.9	0.92	0.7	0.72	0.7	0.65	0.65	0.7	0.83	-0.66	0.34	0.44	0.18	0.74	0.39	0.31	0.75
MDVP:RAP	-0.076	0.097	-0.1	0.99	0.92	1	0.96	1	0.76	0.79	0.74	0.71	0.74	0.74	0.92	-0.72	0.27	0.34	0.064	0.65	0.32	0.43	0.67
MDVP:PPQ	-0.11	0.091	-0.096	0.97	0.9	0.96	1	0.96	0.8	0.84	0.76	0.79	0.8	0.76	0.84	-0.73	0.29	0.33	0.2	0.72	0.41	0.41	0.77
Jitter:DDP	-0.076	0.097	-0.1	0.99	0.92	1	0.96	1	0.76	0.79	0.74	0.71	0.74	0.74	0.92	-0.72	0.27	0.34	0.064	0.65	0.32	0.43	0.67
MDVP:Shimmer	-0.098	0.0023	-0.14	0.77	0.7	0.76	0.8	0.76	1	0.99	0.99	0.98	0.95	0.99	0.72	-0.84	0.37	0.45	0.16	0.65	0.45	0.51	0.69
MDVP:Shimmer(dB)	-0.074	0.043	-0.12	0.8	0.72	0.79	0.84	0.79	0.99	1	0.96	0.97	0.96	0.96	0.74	-0.83	0.35	0.41	0.17	0.65	0.45	0.51	0.7
Shimmer:APQ3	-0.095	-0.0037	-0.15	0.75	0.7	0.74	0.76	0.74	0.99	0.96	1	0.96	0.9	1	0.72	-0.83	0.35	0.44	0.15	0.61	0.4	0.47	0.65
Shimmer:APQ5	-0.071	-0.01	-0.1	0.73	0.65	0.71	0.79	0.71	0.98	0.97	0.96	1	0.95	0.96	0.66	-0.81	0.35	0.4	0.21	0.65	0.46	0.5	0.7
MDVP:APQ	-0.078	0.0049	-0.11	0.76	0.65	0.74	0.8	0.74	0.95	0.96	0.9	0.95	1	0.9	0.69	-0.8	0.36	0.45	0.16	0.67	0.5	0.54	0.72
Shimmer:DDA	-0.095	-0.0037	-0.15	0.75	0.7	0.74	0.76	0.74	0.99	0.96	1	0.96	0.9	1	0.72	-0.83	0.35	0.44	0.15	0.61	0.4	0.47	0.65
NHR	-0.022	0.16	-0.11	0.91	0.83	0.92	0.84	0.92	0.72	0.74	0.72	0.66	0.69	0.72	1	-0.71	0.19	0.37	-0.13	0.54	0.32	0.47	0.55
HNR	0.059	-0.025	0.21	-0.73	-0.66	-0.72	-0.73	-0.72	-0.84	-0.83	-0.83	-0.81	-0.8	-0.83	-0.71	1	-0.36	-0.6	-0.0087	-0.67	-0.43	-0.6	-0.69
status	-0.38	-0.17	-0.38	0.28	0.34	0.27	0.29	0.27	0.37	0.35	0.35	0.35	0.36	0.35	0.19	-0.36	1	0.31	0.23	0.56	0.45	0.34	0.53
RPDE	-0.38	-0.11	-0.4	0.36	0.44	0.34	0.33	0.34	0.45	0.41	0.44	0.4	0.45	0.44	0.37	-0.6	0.31	1	-0.11	0.59	0.48	0.24	0.55
DFA	-0.45	-0.34	-0.05	0.099	0.18	0.064	0.2	0.064	0.16	0.17	0.15	0.21	0.16	0.15	-0.13	-0.0087	0.23	-0.11	1	0.2	0.17	-0.17	0.27
spread1	-0.41	-0.077	-0.39	0.69	0.74	0.65	0.72	0.65	0.65	0.65	0.61	0.65	0.67	0.61	0.54	-0.67	0.56	0.59	0.2	1	0.65	0.5	0.96
spread2	-0.25	-0.003	-0.24	0.39	0.39	0.32	0.41	0.32	0.45	0.45	0.4	0.46	0.5	0.4	0.32	-0.43	0.45	0.48	0.17	0.65	1	0.52	0.64
D2	0.18	0.18	-0.1	0.43	0.31	0.43	0.41	0.43	0.51	0.51	0.47	0.5	0.54	0.47	0.47	-0.6	0.34	0.24	-0.17	0.5	0.52	1	0.48
PPE	-0.37	-0.07	-0.34	0.72	0.75	0.67	0.77	0.67	0.69	0.7	0.65	0.7	0.72	0.65	0.55	-0.69	0.53	0.55	0.27	0.96	0.64	0.48	1

Figure 2: Correlation between features of the dataset

Result Analysis

Gradient Boosting Classifier

1. **Confusion Matrix:** Figure 3 shows that no normal samples were misclassified as PD, whereas 2 PD samples were misclassified as normal. Additionally, 5 normal samples were correctly classified as not PD, and 13 PD samples were correctly classified as PD.
2. **Classification Report:** Table 2 provides the classification report for the Gradient boosting, showing an overall accuracy of 90%. The macro-average metrics, which treat all classes equally, indicate a balanced performance with a precision of 0.86, recall of 0.93, and F1-score of 0.88. The weightedaverage metrics, taking class distribution into account, report a precision, recall, and F1-score of 0.93, 0.90 and 0.90 respectively, reflecting the strong performance of the model.
3. **PR Curve:** Figure 4 shows high precision across most recall values, indicating accurate positive predictions initially. Precision remains close to 1.0 until recall reaches near 1.0. At that point, precision sharply drops, suggesting an increase in false positives when trying to capture all positives. This pattern implies the model is conservative initially but becomes less precise as it maximizes recall. Overall, the model performs well with high precision until it strives for complete recall.

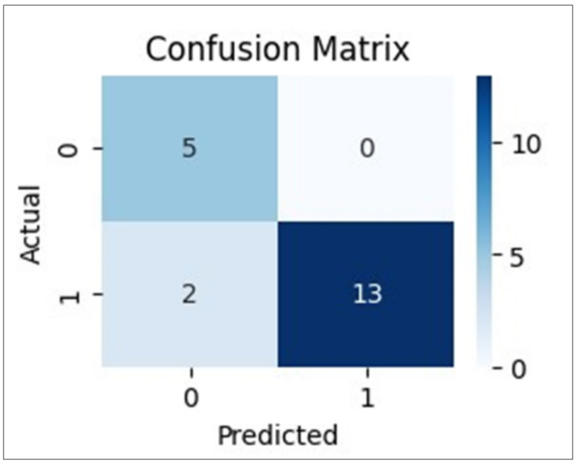


Figure 3: Confusion Matrix for Gradient boosting classifier

Table 2: Classification Report for Gradient Boosting Classifier

	Precision	Recall	F1-Score	Support
0	0.71	1.00	0.83	5
1	1.00	0.87	0.93	15
Accuracy			0.90	20
Macro Avg	0.86	0.93	0.88	20
Weighted Avg	0.93	0.90	0.90	20

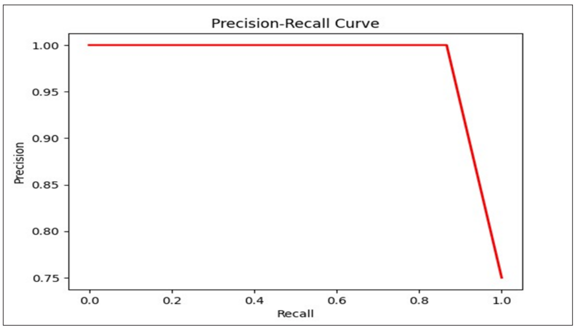


Figure 4: PR Curve for Gradient boosting

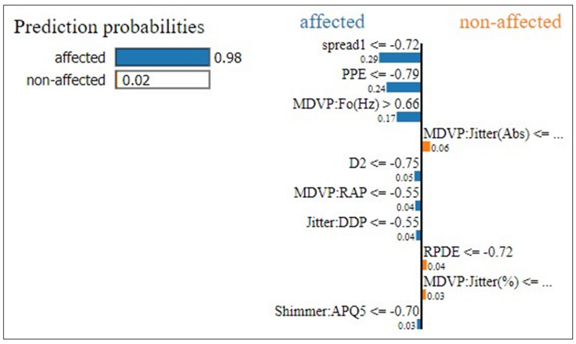


Figure 5: Gradient boosting LIME Tabular plots

LIME: As illustrated in Figure 5, the model predicted Parkinson's disease (affected) for the first test sample with a probability of 98%, as explained by LIME. Key positive features driving this prediction included spread1 $\leq$ -0.72, contributing 0.29, and PPE $\leq$ -0.79, contributing 0.24. Additionally, MDVP:Fo(Hz) > 0.66 supported the affected class with a contribution of 0.17. Conversely, MDVP:Jitter(Abs) $\leq$ ... slightly influenced the prediction toward the non-affected class, with a contribution of 0.06. Other features, such as RPDE $\leq$ -0.72 and D2 $\leq$ -0.75, had a negligible impact on the model's decision.

AdaBoost Classifier

Confusion Matrix: Figure 6 illustrates the classification results, where one normal sample was misclassified as PD, and three PD samples were misclassified as normal. On the other hand, four normal samples were correctly identified as not PD, while 12 PD samples were accurately classified as PD.

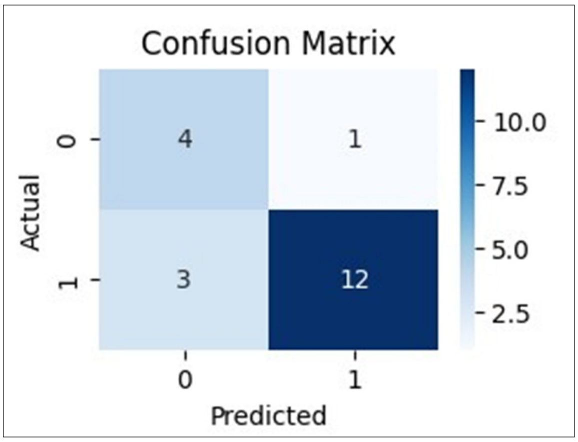


Figure 6: Confusion Matrix for AdaBoost

Classsification Report: Table 3 presents the classification report for the AdaBoost model, which achieved an overall accuracy of 80%. The macro-average metrics, which equally weight all classes, show a precision of 0.75, recall of 0.8, and F1-score of 0.76, indicating balanced performance across classes. Meanwhile, the weighted-average metrics, which account for class distribution, report a precision of 0.84, recall of 0.8, and F1-score of 0.81, highlighting the model's strong overall performance.

PR Curve: Figure 7 demonstrates strong overall performance, starting with high precision and recall, reflecting accurate predictions. However, as recall increases, precision drops noticeably, indicating that the model may face challenges in accurately identifying positive cases when they occur more frequently.

LIME: Figure 8 provides an explanation of the model's prediction for Parkinson's disease, estimating a 67% probability that the individual is affected. Features such as spread1 $\leq$ -0.72 and MDVP $\leq$ -0.66 strongly support the "affected" class. In contrast, features like RPDE $\leq$ -0.72 and Shimmer $\leq$ -0.75 favor the "non-affected" class but carry less influence. Overall, the stronger contributions from features supporting the "affected" class result in a higher prediction probability.

Table 3: Classification Report for AdaBoost

	precision	recall	f1-score	support
0	0.57	0.8	0.67	5
1	0.92	0.8	0.86	15
accuracy			0.8	20
macro avg	0.75	0.8	0.76	20
weighted avg	0.84	0.8	0.81	20

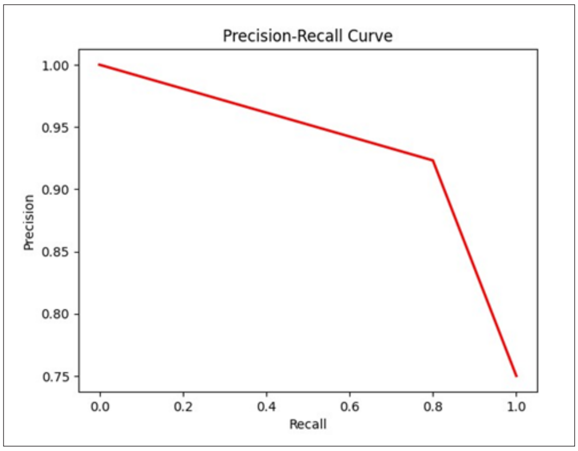


Figure 7: PR Curve for AdaBoost

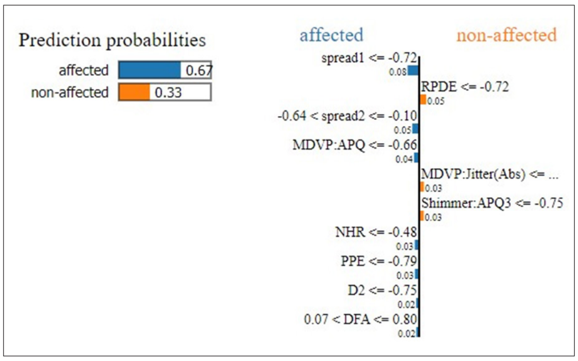


Figure 8: AdaBoost LIME Tabular Plots

KNN Classifier

Confusion Matrix: Figure 9 shows that one PD sample was misclassified as normal. Additionally, five normal samples were accurately identified as not PD, while 14 PD samples were correctly classified as PD.

Classsification Report: Table 4 presents the classification report for the KNN, revealing an overall accuracy of 95%. The macro-average metrics treat all classes equally and demonstrate balanced performance with a precision of 0.92, recall of 0.97, and F1-score of 0.94. The weighted-average metrics, which account for class distribution, report a precision of 0.96, recall of 0.95, and F1-score of 0.95, highlighting the model's robust performance.

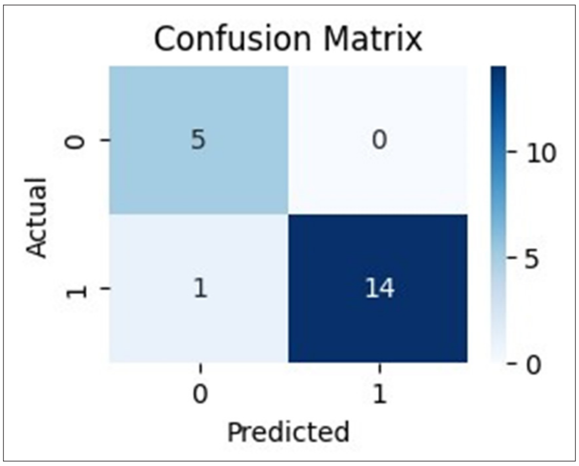


Figure 9: Confusion Matrix for KNN Classifier

Table 4: Classification Report for KNN Classifier

	precision	recall	f1-score	support
0	0.83	1	0.91	5
1	1	0.93	0.97	15
accuracy			0.95	20
macro avg	0.92	0.97	0.94	20
weighted avg	0.96	0.95	0.95	20

PR Curve: Figure 10 shows a similar trend as Adaboost, with a steep drop in precision as recall increases. This suggests that KNN also struggles with maintaining accuracy when classifying more positive cases. Potential improvements include optimizing the number of neighbors (k) and considering feature scaling or dimensionality reduction to enhance the model's performance.

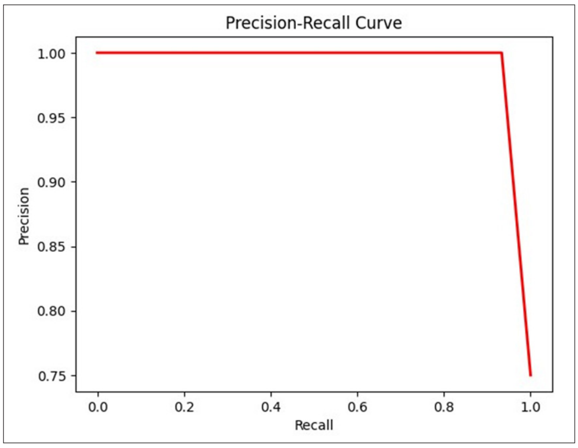


Figure 10: PR Curve for KNN classifier

LIME: Figure 11 reveals the key features influencing the KNN model's prediction for a specific Parkinson's case. ”spread1” and ”MDVP:Fo(Hz)” had the most significant positive impact, increasing the likelihood of an affected diagnosis. Other features like ”PPE” and ”MDVP:Flo(Hz)” had negative contributions. This breakdown provides insights into the model's decision-making process, highlighting the importance of these specific features in Parkinson's disease diagnosis.

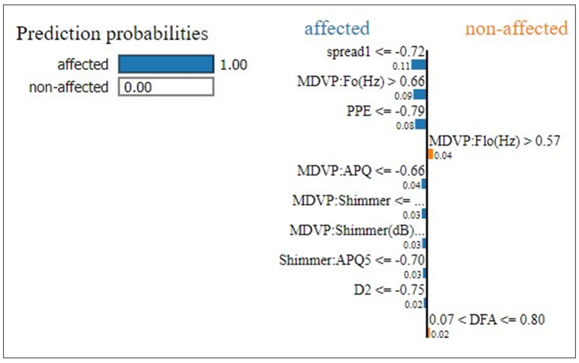


Figure 11: KNN LIME Tabular Plots

LGBM Classifier

Confusion Matrix: Figure 12 shows that 2 PD samples were misclassified as normal. Additionally, 5 normal samples were correctly classified as not PD, and 13 PD samples were correctly classified as PD.

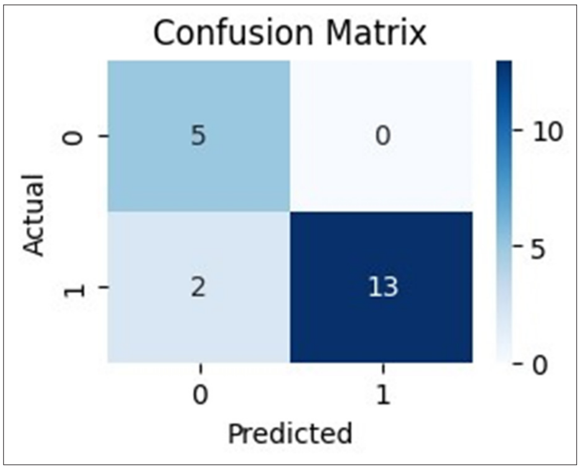


Figure 12: Confusion Matrix for LGBMC

Classsification Report: Table 5 presents the classification metrics for the LGBM model for two classes (0 and 1). Class 1 demonstrates strong performance with a precision of 0.96, recall of 0.99, and F1-score of 0.98, reflecting accurate predictions. In contrast, Class 0 has lower performance with a precision of 0.50, recall of 0.14, and F1-score of 0.22, indicating poor prediction. The model achieves an overall accuracy of 96.78%, with a higher weighted average due to the class imbalance.

PR Curve: Figure 13 shows a steep decline in precision as recall increases. This indicates a trade-off between precision and recall, where the model struggles to maintain accuracy when classifying more positive cases. This might be due to class imbalance or model overfitting.

LIME: Figure 14 shows that ”spread1” and ”PPE” had the most significant positive impact, increasing the likelihood of an affected diagnosis. Other features like ”D2” and ”Shimmer:APQ3” had negative contributions. The model is highly confident in its prediction, with a probability of 0.97 for the ”affected” class. This breakdown provides insights into the model’s decision-making process, highlighting the importance of these specific features in Parkinson’s disease diagnosis.

Table 5: Classification report for LGBM

	precision	recall	f1-score	support
0	0.71	1	0.83	5
1	1	0.87	0.93	15
accuracy			0.9	20
macro avg	0.86	0.93	0.88	20
weighted avg	0.93	0.9	0.9	20

RandomForest Classifier

Confusion Matrix: Figure 15 shows that one normal sample was misclassified as PD, and one PD sample was misclassified as normal. In contrast, four normal samples were accurately identified as not PD, while 14 PD samples were correctly classified as PD.

Classsification Report: Table 6 provides the classification metrics for the random forest model across two classes (0 and 1). Class 1 exhibits strong performance, achieving a precision of 0.93, recall of 0.87, and F1-score of 0.9, indicating accurate predictions. In contrast, Class 0 shows weaker performance, with a precision of 0.67, recall of 0.8, and F1-score of 0.73, reflecting less reliable predictions. The model achieves an overall accuracy of 85.0%, with a higher weighted average driven by class imbalance.

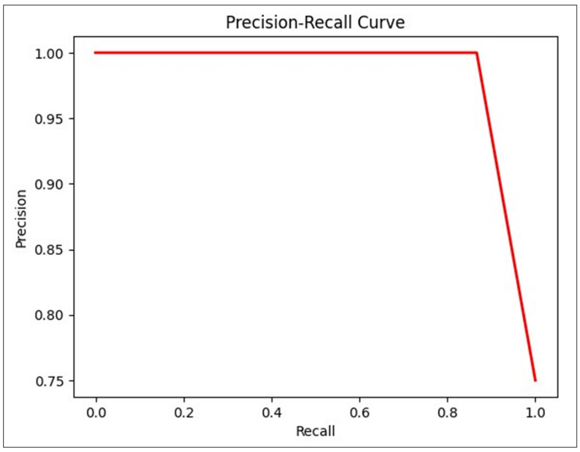


Figure 13: PR Curve for LGBM

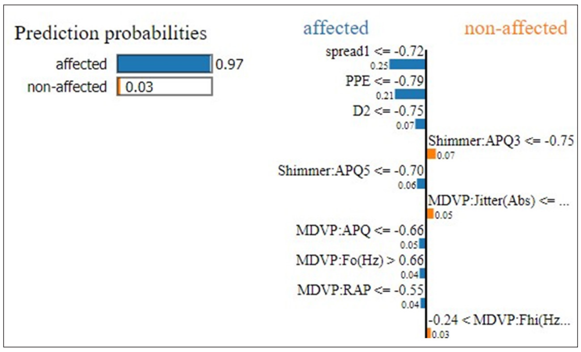


Figure 14: LGBM LIME Tabular Plots

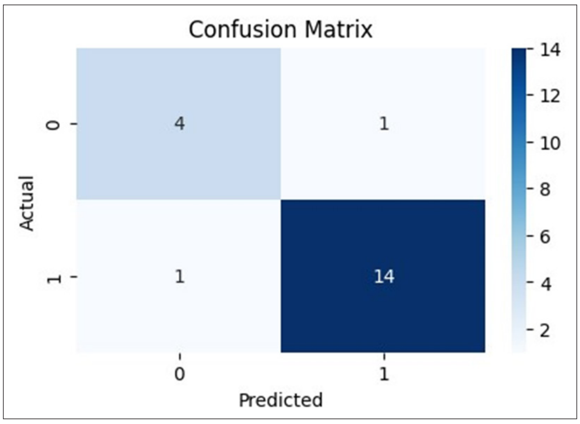


Figure 15: Confusion Matrix for RandomForest

Table 6: Classification report for Random Forest Classifier

	precision	recall	f1-score	support
0	0.67	0.8	0.73	5
1	0.93	0.87	0.9	15
accuracy			0.85	20
macro avg	0.8	0.83	0.81	20
weighted avg	0.86	0.85	0.85	20

PR Curve: Figure 16 illustrates the performance of the Random Forest model. Precision begins at 1.0 when recall is 0.0, demonstrating perfect precision at minimal recall. However, as recall increases, precision gradually decreases, reaching 0.75 at a recall of 1.0, with a noticeable decline after recall exceeds 0.8. This highlights the trade-off between precision and recall, where achieving higher recall comes at the cost of reduced precision. The model maintains high precision at lower recall levels but sacrifices precision as it identifies more positive cases.

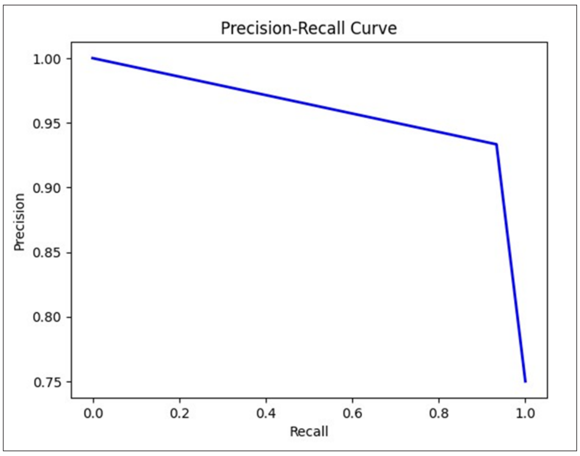


Figure 16: PR Curve for Random Forest

LIME: Figure 17 illustrates that ”spread1”, ”PPE”, and ”MDVP:Fo(Hz)” are the most significant characteristics pushing the Random Forest model’s prediction to the ”affected” class. The low contribution of opposing features (such as ”DFA”) is consistent with the high confidence (88%) in predicting the diagnosis as ”affected.” This research study demonstrates how the Random Forest model analyzes and evaluates certain features while diagnosing Parkinson’s disease.

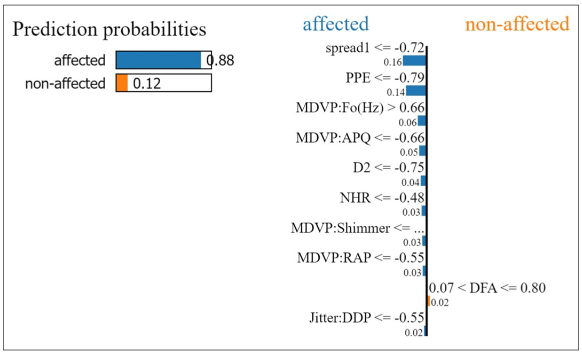


Figure 17: Random Forest LIME Tabular Plots

Comparative Analysis of Statistical Results

Adaboost and Random Forest demonstrated strong performance across multiple metrics, with Adaboost achieving an average training accuracy of 0.90 and Random Forest slightly outperforming it with an average accuracy of 0.914. Both models achieved high accuracy in most subsets, particularly in k1, where each reached 0.97. This consistency underscores their ability to generalize effectively across the training data, making them reliable choices for high-accuracy classification tasks.

LightGBM (LGBM) showed comparable performance, boasting an average accuracy of 0.93. It maintained strong accuracy across all subsets, with a minimum of 0.82 in k4 but otherwise reaching 0.94 in multiple instances. LGBM’s stability highlights its robustness and suitability for tasks requiring consistent performance across varied data distributions. K-Nearest Neighbour (KNeighbour) also performed admirably, with an average accuracy of 0.90. Although its accuracy slightly fluctuated, ranging from 0.82 in k4 to 0.94 in k3 and k5, it consistently produced solid results. KNeighbour provides a balanced choice between accuracy and consistency, making it suitable for applications where moderate variation across subsets is acceptable.

Gradient Boosting exhibited some variability, achieving an average accuracy of 0.89. It peaked at 0.97 in k2 but fell to 0.85 in k3 and k5, indicating a tendency to overfit certain data splits. This variability suggests that, while effective, Gradient Boosting may benefit from further tuning to improve stability. The results (Table 7) underscore that Adaboost, LGBM, Random Forest, and KNeighbour are well-suited for tasks requiring high accuracy, with LGBM and Random Forest showing a slight edge in stability and overall performance.

Table 7: Comparative Analysis of Training Accuracy

Fold	Adaboost	Gradient B.	KNN	LGBM	Random F.
k1	0.97	0.91	0.91	0.97	0.97
k2	0.85	0.97	0.91	0.94	0.91
k3	0.88	0.85	0.94	0.94	0.94
k4	0.88	0.88	0.82	0.85	0.85
k5	0.91	0.85	0.94	0.94	0.88
Average	0.9	0.89	0.9	0.93	0.914

Gradient Boosting, while competitive, may require additional adjustments to achieve consistency across diverse datasets [20].

Conclusion

This study employed various machine learning algorithms to classify Parkinson’s disease (PD) and non-PD cases. Among the models evaluated, the LGBMClassifier emerged as the most accurate and reliable, achieving an average accuracy of 93.00%, a precision of 0.90, a recall of 0.87, and an F1-score of 0.93 for the ”PD” class. Its balanced performance across both classes highlights its potential for practical application in medical diagnostics. The model’s ability to accurately detect PD cases while minimizing misclassification of non-PD cases underscores its suitability for early detection and timely intervention. These results indicate that the LGBMClassifier could serve as a valuable tool for enhancing diagnostic accuracy in Parkinson’s disease. Additionally, further improvements could be achieved by applying optimization techniques [21,22] and implementing strategies to address adversarial challenges in neural networks [23,24].

References

1. Bloem RB, Okun MS, Klein C. Parkinson's disease. The Lancet. 2021. 397: 2284-2303.
2. Gaur L, Bhandari M, Razdan T, Mallik S, Zhao Z. Explanation-driven deep learning model for prediction of brain tumour status using MRI image data. Frontiers in genetics. 2021. 13: 822666.
3. Bhandari M, Shahi TB, Neupane A, Walsh KB. Botanicx-ai: Identification of tomato leaf diseases using an explanation-driven deeplearning model. Journal of Imaging. 2023. 9: 53.
4. Bhandari M, Shahi TB, Siku B, Neupane A. Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI. Computers in Biology and Medicine. 2022. 150: 106156.
5. Bhandari M, Yogarajah P, Kavitha MS, Condell J. Exploring the capabilities of a lightweight CNN model in accurately identifying renal abnormalities: Cysts, stones, and tumors, using LIME and SHAP. Applied Science. 2023. 13: 3125.
6. Gaur L, Bhandari M, Bhadwal SS, Jhanjhi N, Mohammad S. et al. Explanation-driven HCI model to examine the mini-mental state for Alzheimer's disease. ACM Transactions on Multimedia Computing, Communications and Applications. 2023. 20: 1-16.
7. 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.
8. 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.
9. 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 (SCOPES). IEEE. 2016. 1446-1451.
10. Yadav S, Singh MK, Pal S. Artificial intelligence model for parkinson disease detection using machine learning algorithms. Biomedical Materials & Devices. 2023. 1: 899-911.
11. Mittal V, Sharma RK. Machine learning approach for classification of Parkinson disease using acoustic features. Journal of Reliable Intelligent Environments. 2021. 7: 233-239.
12. Shivangi, Johri A, Tripathi A. Parkinson disease detection using deep neural networks. 2019 Twelfth international conference on contemporary computing (IC3). IEEE. 2019. 1-4.
13. Wang W, Lee J, Harrou F, Sun Y. Early detection of Parkinson's disease using deep learning and machine learning. IEEE Access. 2020 8: 147635-147646.
14. Bhandari M, Neupane A, Mallik S, Gaur L, Qin H. Auguring fake face images using dual input convolution neural network. Journal of Imaging. 2022. 9: 3.
15. Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Nature Precedings. 2007. 1.
16. Jebreili S, Goli A. Optimization and computing using intelligent data-driven. Optimization and Computing using Intelligent Data-Driven Approaches for Decision-Making: Optimization Applications. 2024. 90.
17. Khanal M, Khadka SR, Subedi H, Chaulagain IP, Regmi LN. et al. Explaining the factors affecting customer satisfaction at the fintech firm F1 soft by using PCA and XAI. FinTech. 2023. 2: 70-84.
18. Dowling A, O'Dwyer J, Adley CC. LIME in the LIMElight. Journal of cleaner production. 2015. 92: 13-22.
19. Garreau D, Luxburg U. Explaining the explainer: A first theoretical analysis of LIME. International conference on artificial intelligence and statistics. PMLR. 2020. 1287-1296.
20. Bentéjac C, Csörgő A, Muñoz GM. A comparative analysis of gradient boosting˜ algorithms. Artificial Intelligence Review. 2021. 54: 1937-1967.
21. Bhandari M, Panday S, Bhatta CP, Panday SP. Image Steganography Approach Based Ant Colony Optimization with Triangular Chaotic Map. 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM). 2022. 2: 429-434.
22. 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.
23. Bhandari M, Shahi TB, Neupane A. Evaluating Retinal Disease Diagnosis with an Interpretable Lightweight CNN Model Resistant to Adversarial Attacks. Journal of Imaging. 2023. 9: 219.
24. Gaur L, Bhandari M, Razdan T. Development of Image Translating Model to Counter Adversarial Attacks. DeepFakes. CRC Press. 2022. 57-72.