Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 20s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Predictive Modelling of Parkinson's Disease Progression Using LightGBM Classifier

Jyotsnarani Tripathy[1], D. Sowjanya[2], Chunduru Anilkumar[3], Subramanyam Kunisetti [4], K. Chandramouli[5]

[1]Department of CSE (AIML & IoT), VNR Vignan Jyothi Institute of Engineering and Technology, Nizampet, Hyderabad, T.S, India

[2]Department of CSE, GMR Institute of Technology, Rajam, Andhra Pradesh, India

[3]Department of Computer Science Engineering (Data Science), Aditya Institute of Technology and Management, Tekkali, Srikakulam, Andhra Pradesh 532201

[4]Department of CSE (Data Science), R.V.R & J.C College of Engineering, Chowdavaram,

Guntur, Andhra Pradesh, India.

[5]Department of Information Technology, Aditya Institute of Technology and Management,

Tekkali, Srikakulam, Andhra Pradesh 532201

***Corresponding Author:**

Email ID: anilchunduru.csd@adityatekkali.edu.in

Cite this paper as: Jyotsnarani Tripathy, D. Sowjanya, Chunduru Anilkumar, Subramanyam Kunisetti , K. Chandramouli, (2025) Predictive Modelling of Parkinson's Disease Progression Using LightGBM Classifier. *Journal of Neonatal Surgery*, 14 (20s), 279-292.

ABSTRACT

Parkinson's disease is a progressive neurological disorder that affects millions worldwide, causing a variety of motor and non-motor symptoms, such as the condition known as freezing of gait (FOG). The overall health of people with Parkinson's disease can be significantly enhanced by early identification and treatment of freezing of gait (FOG). Using LightGBM (Gradient Boosting Machine), a well-liked gradient boosting ensemble method that is trained using the AutoML tool and is based on decision trees, we present a prediction model for freezing of gait in this work. Leveraging a comprehensive dataset of Parkinson's disease patient's clinical profiles, gait patterns, and demographic information, we employed feature engineering techniques to extract meaningful predictors associated with FOG. Our results demonstrate the effectiveness of the LightGBM model in accurately predicting FOG episodes in Parkinson's patients. The model evaluation shown that LightGBM offers the best results with an accuracy of 92.31% when compared to other models like Support Vector Classifier and Logistic Regression.

Keywords: *Freezing of gait (FOG), LightGBM, Motor and non-motor symptoms, Gait patterns, Feature engineering, Gradient boosting*

1. INTRODUCTION

Parkinson's disease includes the slow degeneration or death of particular neuronal cells in the brain and is thought to be the second most common age-related neurodegenerative disorder, affecting over 10 million people globally. One of the key indicators of Parkinson's disease is the degeneration of neurons responsible for producing dopamine. Dopamine deficiency results in disturbed brain function and a variety of symptoms, including trouble moving. Among these symptoms, "freezing of gait" (FOG), which is defined by sudden, transient episodes of being unable to commence or continue walking and results in momentary moments of being trappedin place, is a common and distressing one. FOG can also affect other movements, such as difficulty starting or stopping when turning or navigating through narrow spaces and common symptoms are tremors (shaking of hands, fingers or other body parts),muscle rigidity, shuffling,short stepped gait, fatigue, speech and swallowing difficulties, sleep disturbances.

Parkinson's disease is expected to kill 329,000 individuals worldwide in 2022. This numberhas increased now for a number of reasons, included a population that is aging and a rise in Parkinson's disease diagnoses. An estimated 60,000 Americans every year in the USA pass away from the neurological disorder Parkinson's. By 2030, this figure is predicted to rise to 100,000. Men die at a higher rate than women from Parkinson's disease. This is most likely because men are initially diagnosed with the disorder at a higher rate than women. Parkinson's disease patients typically pass away at age 75. In 1967, Hoehn and Yahr established five stages of the disorder based on clinical issues. Professionals apply a classification scheme to characterize how Parkinson's disease's motor symptoms evolve. Parkinson's disease has several phases, with stages 1 and

2 indicating the early stages, stages 3 and 4 the middle stages, and stages 4 and 5 the later stages.

A person goes through this stage with moderate symptoms that do not disrupt daily activities. The only part of the body where tremor and other movement symptoms appear is that side. Changes are made to one's movement, expressions, and posture. The symptoms worsen over time, eventually affecting the midline (such as the neck and trunk) or both sides of the body. Movement anomalies including tremors, rigidity, and others become more obvious. Thoseaffected might have bad posture and have trouble walking. During this medium stage, they are still capable of living alone, but daily duties grow more difficult and time-consuming. Loss of balance becomes obvious, which causes instability when turning or being propelled from a standing position, and an increase in the number of falls. The individual's daily activities are increasingly limited by their worsening motor symptoms, which range in severity from mild handicap to severe impairment. The symptoms are now at their most severe and disable phase. The person may use a cane or walker for safety even though they are still able to stand and walk unaided. However, they can no longer live independently and need a lot of assistance with their everyday activities.

The most serious and perhaps fatal stage is this one. Severe leg stiffness makes it difficult to move around. The person may use a cane or walker for safety even though they are still able to stand and walk on their own. They require a lot of help with their daily chores, making living alone impractical.

The continuing loss of brain cells and neurons in the substantia nigra region of the brain is responsible for the patients' gradually deteriorating motor function. Rare genetic types of Parkinson's disease with early appearance have been connected to synuclein protein mutations. The development and movement of Parkinson's disease may be significantly influenced by the presence of aberrant clusters (aggregates) of the same protein in functioning neurons within the affected brain areas. The particular causes of the troublesome synuclein clumping, which results in the signs and symptoms of Parkinson's disease, are yet unknown

2. RELATED WORK:

The Daphnet dataset was utilized in this study to compare the results reported by the authors to the most recent work done by other researchers who also used the same dataset. The potential applications of this research include the development of wearable devices that can detect and monitor FOG episodes in real-time, which can help patients to avoid falls. Wearable technology is accessible, small, and has a long battery life; all users need to do to utilize them is attach them to their bodies and turn them on. In terms of sensitivity and specificity, the patient-dependent model created in this study fared better than prior FOG detection techniques. The model outperformed results from earlier studies that used the same dataset, achieving a sensitivity of 92.5% and a specificity of 95.6% on the test set.[1]

The unsupervised convolutional denoising autoencoder model works by first pre-processing the gait data from wearable sensors to remove noise. Then, the model learns to extract features from the pre-processed data using unsupervised learning techniques. These learned features are then used to classify gait patterns as either normal or freezing of gait. To get the best results, the model is adjusted. First off, it does away with the requirement for manually handcrafting features, which takes time and makes choosing the best features more challenging. Second, by employing convolution and pooling processes, the model may automatically discover feature representations of the data. Thirdly, the model might curtail or do away with the use of characteristics that are made by hand. Finally, the model can use wearable technology to continuously and accurately analyze the gait of Parkinson's disease patients.[2]

Early Parkinson's disorder detection is important because it allows for immediate support to slow disease progression and lower patient morbidity. In order to identify gait characteristics objectively, which is crucial for treating patients with Parkinson's disorder as well as patients with varying degrees of disease severity, neural network models can be of assistance in this respect. The study in this file used convolutional neural networks and linear discriminant analysis to categorize Parkinson's disorder and its phases in 54 participants with an accuracy of up to 90.62%. According to, this study has significant consequences for how Parkinson's condition will be identified and treated in the future. This might slow the spread of the illness and lower the mortality rate for patients.[3]

The authors reviewed a number of studies that have used different types of data, including clinical features, imaging data, and speech data, to train machine learning models to predict the presence of Parkinson's disorder. The study took consideration of thirteen signs based on data from the Parkinson's Progression Markers Initiative (PPMI) dataset to identify early Parkinson's disorder. These characteristics include demographic data, physical symptoms, and psychological symptoms like depression, anxiety, and sleep problems .The performance of individual models is significantly improved by the ensemble network, which aggregates the output of three deep learning networks. The comparison demonstrated the developed model's improved detection performance, which averages the greatest accuracy at 96.45%.[4]

The proposed method in the paper is a spatial-temporal graphical convolution network (ST-GCN). Time series data can be utilized to learn the temporal and spatial interactions in ST-GCNs, a sort of deep learning model. The authors of the paper trained an ST- GCN on kinematic data from Parkinson patients, and they were able to achieve an accuracy of 97.6% in

scoring FOG. The study's use of a relatively small sample of patients is another drawback of the paper. Overall, the study has positive results and suggests a fresh approach to evaluating FOG. The ST-GCN model was able to achieve high accuracy in scoring FOG, and it is a 4.promising tool for the objective and reliable assessment of FOG in Parkinson patients.[5]

The paper begins by providing an overview of FOG and its impact on people with PD. It then discusses the challenges of detecting FOG, including the variability of the symptom and the need for energy-efficient algorithms. The paper also examines the various contexts, such as the user's environment, their current activity, and their previous gait patterns, that can be used to enhance the accuracy of FOG detection. Overall, the study offers a useful summary of the research on context recognition algorithms for FOG detection. Future efforts to create more precise and energy-efficient FOG detection systems are likely to heavily rely on context recognition algorithms.[6]

This paper proposes a novel method for analyzing and predicting Parkinson's disease (PD) data. The proposed method, called MV-SAE, is a multi-variant stacked auto encoder that uses multiple features to capture the complex patterns of PD data. The authors evaluated the performance of MV-SAE on the UCI Parkinson's Disease Data Set, which consists of 20 features from 195 PD patients and 195 healthy controls. MV-SAE achieved an accuracy of 90.8%, which outperformed the baseline methods, including SVM, Random Forest, and KNN. The authors conclude that MV-SAE is an effective method for analysing and predicting PD data. The method is able to capture the complex patterns of PD data, and it outperforms the baseline methods. The authors suggest that MV-SAE could be used for clinical applications, such as early diagnosis and treatment of PD.[7]

This study proposes a novel method for recognizing Parkinsonian gait from forward videos. The suggested method, WM-STGCN, makes use of a weighted adjacency matrix with virtual connections and multi-scale temporal convolution. A spatiotemporal graph convolutional network is used in this technique. This method aims to record both the spatial and temporal aspects of gait. The UPen-3D gait dataset, which contains gait data from 60 Parkinson's patients and 60 healthy people, was used by the scientists to evaluate WM-STGCN's performance. WM-STGCN outscored conventional techniques such as LSTM, KNN, decision trees, AdaBoost, and ST-GCN, with an F1 score of 92.85%. In addition, it showed a rate of accuracy of 87.1%[8]

In this study implemented artificial neural networks (ANN), decision trees, and support vector machines (SVM) as machine learning techniques. The system outperformed previous techniques in comparison to evaluations done on a dataset that included both Parkinson's disease patients and healthy controls. The proposed Relief feature selection method using Bacterial Foraging Optimization (RF-BFO-SVM) outperforms advanced machine learning techniques like Particle Swarm Optimization (PSO-SVM), Grid-SVM, Kernel Extreme Learning Machine (KELM), and Random Forest (RF), and also delivers more reliable and consistent results in classification tasks. With a classification accuracy of 97.42% in this research, the proposed framework performed quite well.[9]

In this study, the author applied sensor-based datasets that included behavioral signals (acceleration, force, pressure, etc.) from human body motion that were assessed in order to carry out efficient gait analysis. This article discusses a variety of computer vision methods for diagnosing Parkinson's disease, including wearable sensors, RGB and depth cameras, and motion capture with and without markers. The authors also cover feature selection and gait analysis using machine learning algorithms. Finally, the author came to the conclusion that, when it comes to VB, the marker-less technique has been selected and can offer a deeper assessment of PD-affected patients.[10]

3. METHODS:

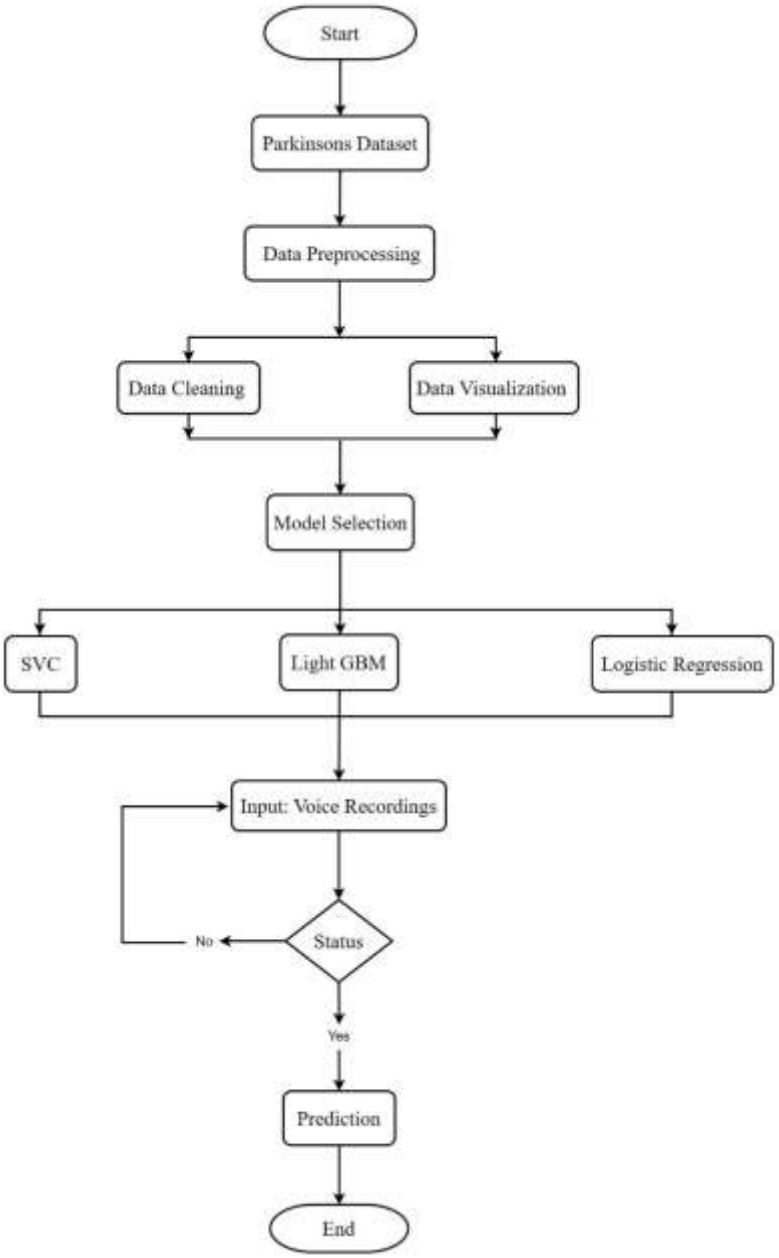


Figure 3.1: Workflow of the Model

3.1 Problem Definition:

A neurodegenerative disorder called Parkinson's disease is marked by signs like postural instability, frozen posture, and resting tremors. The detection of Parkinson's disease is accomplished using a variety of methods, such as LightGBM, Convolutional Neural Networks, Logistic Regression, and Support Vector Machine.

3.2 Data Collection:

The voice signals were captured and compiled into the dataset by Max Little of the University of Oxford. The feature extraction techniques for general voice problems were published in this work.

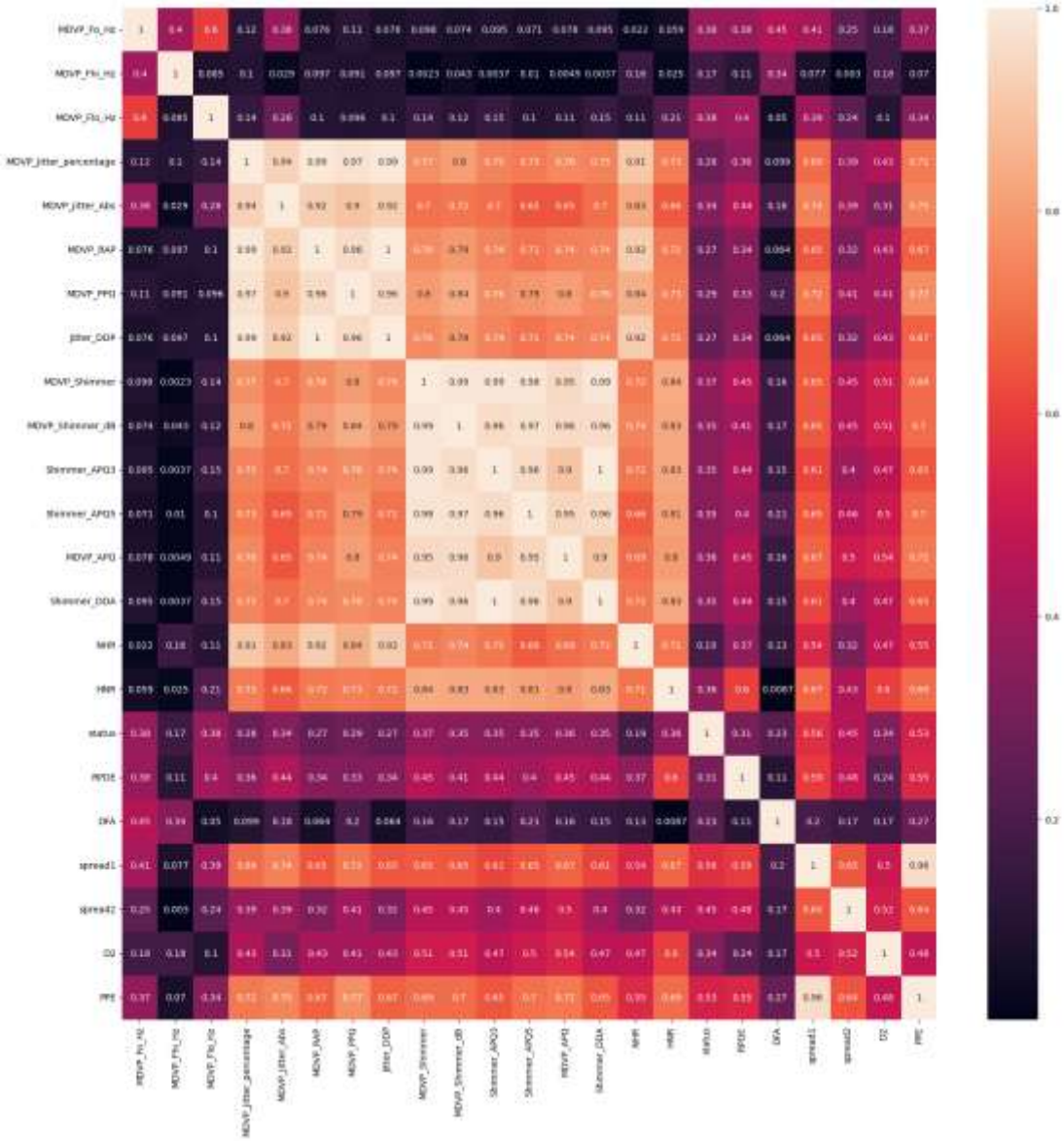
This dataset contains several biological voice measurements taken from 31 individuals, 23 of whom had Parkinson's disease (PD). Each row in the table represents one of the 195 voice recordings from these people, and each column represents a specific voice measure (the "name" column). According to the "status" column, which is set to 0 for healthy and 1 for PD, the primary goal of the data is to distinguish between healthy individuals and those with PD.

	name	MDVP_Fo_Hz	MDVP_Fhi_Hz	MDVP_Flo_Hz	MDVP_Jitter_percentage	MDVP_Jitter_Abs	MDVP_RAP	MDVP_PPQ	Jitter_DDP	MDVP_Shimmer	Shimmer_DDA	NHR	HNR	status	RPDE	DFA	spread1	spread2	D2	PPE
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	0.00554	0.01109	0.04374	0.06545	0.02211	21.033	1	0.414783	0.815285	-4.813031	0.266482	2.301442	0.284654
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	0.09403	0.01929	19.085	1	0.458359	0.819521	-4.075192	0.335590	2.486855	0.368674
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	0.00781	0.01633	0.05233	0.08270	0.01309	20.651	1	0.429895	0.825288	-4.443179	0.311173	2.342259	0.332634
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	0.08771	0.01353	20.644	1	0.434969	0.819235	-4.117501	0.334147	2.405554	0.368975
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	0.10470	0.01767	19.649	1	0.417356	0.823484	-3.747787	0.234513	2.332180	0.410335

Figure 3.2: This dataset consists of 195 records and 24 attributes based on the patient's vocal frequency.

3.3 Data Preprocessing:

In data mining, cleaning includes converting unstructured data into a format that can be understood. Real-world data is frequently rife with gaps, contradictions, missing patterns, and various errors. This may lead to the capture of poor-quality data, which will ultimately influence the caliber of models built using it. An essential first step in overcoming these difficulties is data cleaning.



Preprocessing data is frequently necessary for machine learning methods in order to change or encode it in a way that enables more efficient processing by the computer, making the aspects of the data more accessible and intelligible to thealgorithms.

Figure 3.3 : Correlation between the attributes

3.4 Data Cleaning:

Real-world data frequently lacks accuracy, contains mistakes, and is inconsistent. It can be lacking or contain unrelated information. Data cleansing techniques are used to address these problems. By filling in missing numbers, decreasing noise, identifying outliers, and removing undesirable variations, these strategies seek to correct data discrepancies. Data that is unclear might be confusing for both the model and the data. Therefore, using various data cleaning procedures to assure data quality is a vital stage in the data preprocessing process.

3.5 Checking for duplicates

If the same row or column appears more than once, you can remove the duplicates while keeping the first instance. To avoid giving a particular data object a benefit or bias while running machine learning algorithms.

3.6 Estimate missing values

Simple interpolation techniques can be used to close the gaps if only a small portion of the values are missing. The mean, median, or mode value for an attribute is usually utilized to fill in the gaps left by missing data.

3.7 Data Visualization

The visual presentation of data in the form of images, bar graphs, pie charts, info-graphics, and other visual representations is known as data visualization. In addition to giving you an understanding of statistics, visualizing data enables you to make an accurate conclusion and convey it to others.

Data analysis and big data projects have become increasingly popular, which has highlighted the value of data visualization. The effort of sorting through, understanding, and explaining this data becomes difficult and time-consuming when businesses use machine learning to gather large datasets.

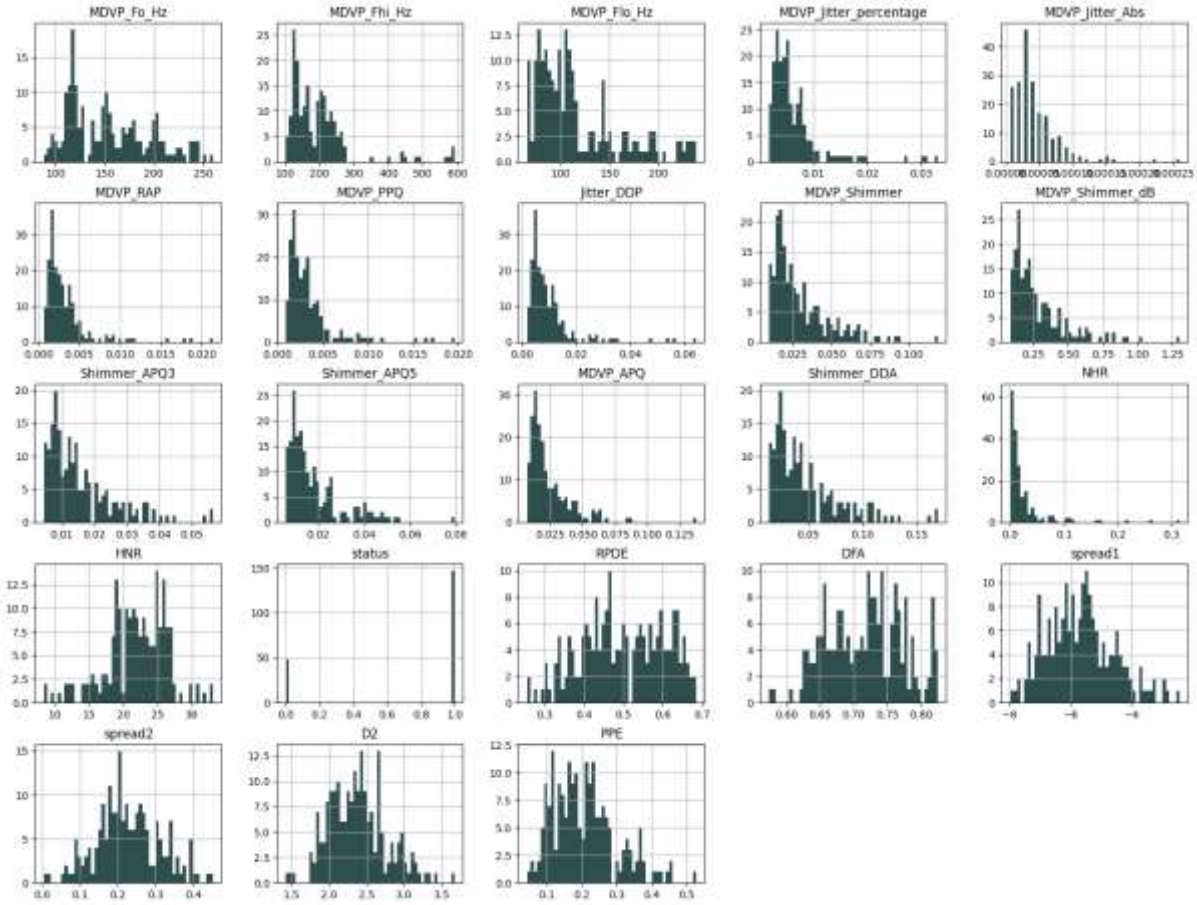


Figure 3.4: Representation of the frequencies of each attribute

4. MODEL SELECTION:

It is the phase in which we select the best and precise model that best fits the problem definition. The process of choosing which method and model architecture is best suited for a specific job or dataset is known as model selection. It comprises contrasting multiple models, evaluating their effectiveness, and selecting the one that best resolves the current situation.

4.1 LightGBM:

LightGBM employs a leaf-wise tree development strategy, in which the algorithm selects the leaf (also referred to as the terminal node) that would cause the greatest reduction in the loss function at each step of tree formation. In other words, it chooses the split that results in the greatest improvement in projected accuracy. LightGBM divides the tree leaf-wise whereas other boosting methods develop the tree level-wise. The leaf with the largest delta loss is chosen to grow.

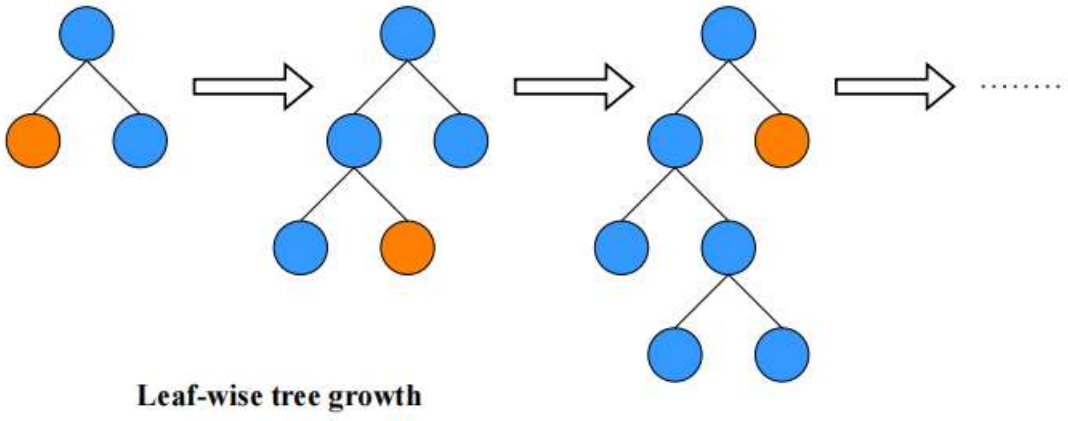


Figure 4.1 : Architecture of Leaf-wise Tree Growth

Best-first (leaf-wise) and depth-first (level-wise) will produce the same tree if we develop the entire tree. As the tree is expanded, there is a variation in the order. Order is important because trees aren't typically grown to their full depth.

Different trees can be produced by using different trimming techniques and early termination criteria. Leaf-wise will frequently (but not always) learn lower-error trees "faster" than level- wise because it chooses splits based on their contribution to the overall loss rather than just the loss along a specific branch.

Leaf-wise will probably perform better than level-wise for a small number of nodes. Because they will eventually physically build the same tree, as we add more nodes they will converge to the identical performance without pausing or trimming.

4.2 Support Vector Classifier:

A supervised machine learning approach used for binary and multi-class classification tasks is a support vector classifier (SVC), commonly referred to as a support vector machine (SVM) for classification. It is a strong and flexible method that excels at working with high-dimensional feature spaces and complex datasets. Finding a hyperplane that best divides data points of distinct classes while increasing the margin between them is the main goal of SVMs.

4.3 Logistic Regression:

The main application of logistic regression is to assess the likelihood that a given instance belongs to a given class in classification problems. Logistic regression uses a sigmoid function to predict the likelihood of an instance belonging to a certain class based on the outcome of a linear regression function, in contrast to linear regression, which produces continuous output. In essence, logistic regression makes predictions about the likelihood of being in a certain class, whereas linear regression generates results with no restrictions on the number of variables.

5. MODEL EVOLUTION:

It is the crucial stage in determining the model's efficiency. By using metrics such as accuracy, precision, recall, and F1 score, cross validation, It is possible to learn more about the model's advantages and disadvantages.

5.1 Accuracy

In classification tasks, accuracy is a commonly used metric that measures how accurate a model's predictions are by comparing the proportion of correct predictions to all guesses. Although accuracy is a simple metric, it may not be the best option for datasets with a notable class imbalance, in which one category significantly outnumbers the others in terms of data

points.

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

5.2 Precision

Precision is a metric that quantifies how many of the positive predictions made by a model were correct. It is calculated as the ratio of true positives (correctly predicted positive instances) to the total number of positive predictions. Precision is essential when minimizing false positives is crucial.

Precision = $\frac{TP}{TP+FP}$

5.3 Recall

The capacity of a model to accurately identify every pertinent instance within a dataset is measured by recall. It is determined as the proportion of actual positive instances to all true positive instances. When minimizing false negatives is essential, recall is key.

Recall = $\frac{TP}{TP+FN}$

5.4 F1 Score

The F1 Score is the harmonic mean of precision and recall. It balances precision and recall, making it a useful metric when you want to strike a balance between false positives and false negatives. It is especially valuable when dealing with imbalanced datasets.

F1-Score = $\frac{2.Precision.Recall}{Precision+Recall}$

5.5 Cross-Validation

Cross-validation is a technique for assessing the performance of a model that involves segmenting the dataset into many "folds" (such as 5, 10, or 20 folds). The performance metrics that arise from training and testing the model on several subsets in each fold are averaged to give a more thorough evaluation of its capabilities. Cross-validation helps estimate how well a model will perform on unseen data and reduces the risk of over fitting. Common types include k-fold cross-validation and stratified cross-validation.

6. DATABASE CONNECTIVITY:

We used the Streamlit framework for our project's deployment and database connection. A Python module called Streamlit has become extremely well-known for its ease-of-use in converting data scripts into shareable web apps. Deploying these applications, however, may occasionally be difficult, particularly when it comes to controlling server infrastructure and scaling for user demand. We now have a basic understanding of streamlit clouds. Streamlit offers a platform called Streamlit Cloud that makes it easier to share and deploy Streamlit apps. Without the effort of setting up servers, managing dependencies, and navigating deployment complexity, it provides an infrastructure to host your apps. This allows developers and data scientists to concentrate on creating the app itself while Streamlit Cloud handles the deployment procedure.

For installing the streamlit we use the following command in command prompt.

```
PS C:\Users\Jagadeesh\OneDrive\Desktop\Project\Parkinson-disease-detection-main\Parkinson-disease-de
tection-main> pip install streamlit
Requirement already satisfied: streamlit in c:\users\jagadeesh\appdata\local\packages\pythonsoftware
foundation.python.3.10_qbz5n2kfra8p0\localcache\local-packages\python310\site-packages (1.26.0)
Requirement already satisfied: altair<6,>=4.0 in c:\users\jagadeesh\appdata\local\packages\pythonsof
twarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-packages\python310\site-packages (from st
reamlit) (5.1.1)
Requirement already satisfied: blinker<2,>=1.0.0 in c:\users\jagadeesh\appdata\local\packages\python
softwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-packages\python310\site-packages (from
 streamlit) (1.6.2)
Requirement already satisfied: cachetools<6,>=4.0 in c:\users\jagadeesh\appdata\local\packages\pytho
nsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-packages\python310\site-packages (fro
m streamlit) (5.3.1)
```

Figure 6.1 : Installing Streamlit

once streamlit installed successfully, then you need to run your code by using the following command in the command prompt.

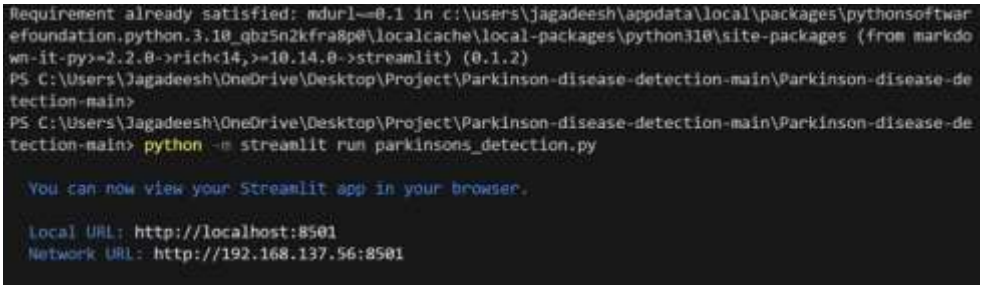


Figure 6.2 : Running Streamlit

7. DEPLOYMENT:

It refers to the process of making the Prediction of Parkinson's disease model accessible and operational for use by end-users. A repository object store is created to store and categorize documents after deployment, and a project can be connected to a user application for document processing.

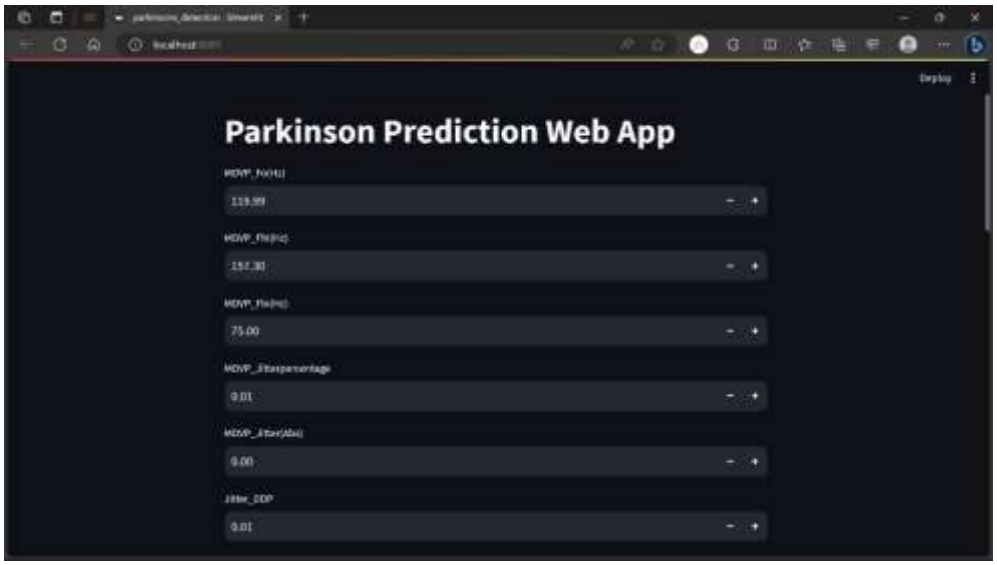


Figure 7.1 : Parkinson's Disease Prediction Web Page

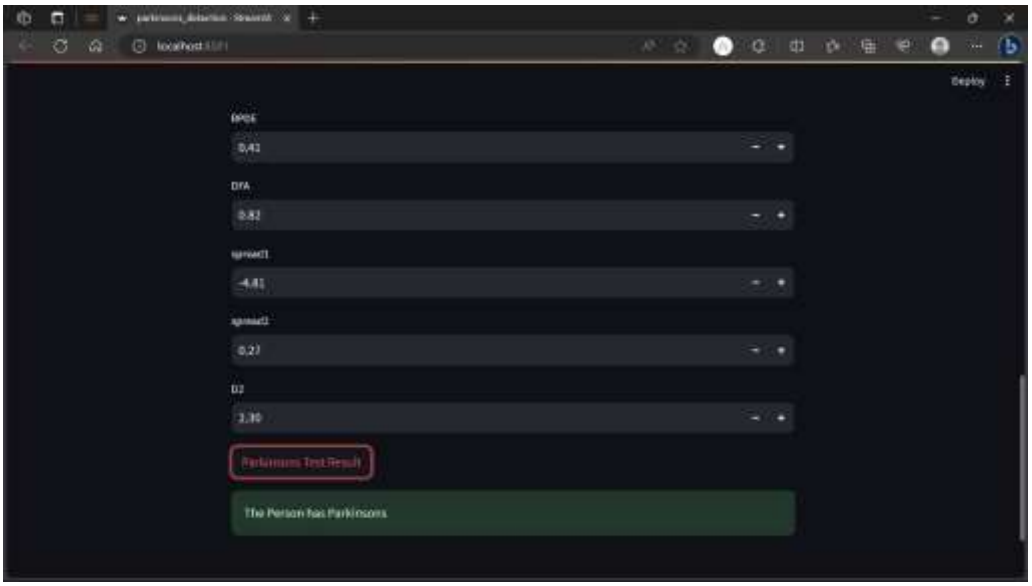


Figure 7.2: Parkinson's Disease Prediction Results Page

8. RESULTS AND DISCUSSION:

This table shows the different evolution metrics obtained for various algorithms and the testing accuracy of LightGBM Algorithm is greater among all the models that is 92.31%.

Table 8.1 : Performance of Models

Evolution	Light GBM	Support Vector Classification	Logistic Regression
Training Accuracy Score	100	84.62	82.69
Cross Validation Score	89.21	84.71	82.17
Testing Accuracy Score	92.31	87.18	87.18
Precision Score	93.94	90.91	90.91
Recall Score	96.88	93.75	93.75
F1 Score	95.38	92.31	92.31

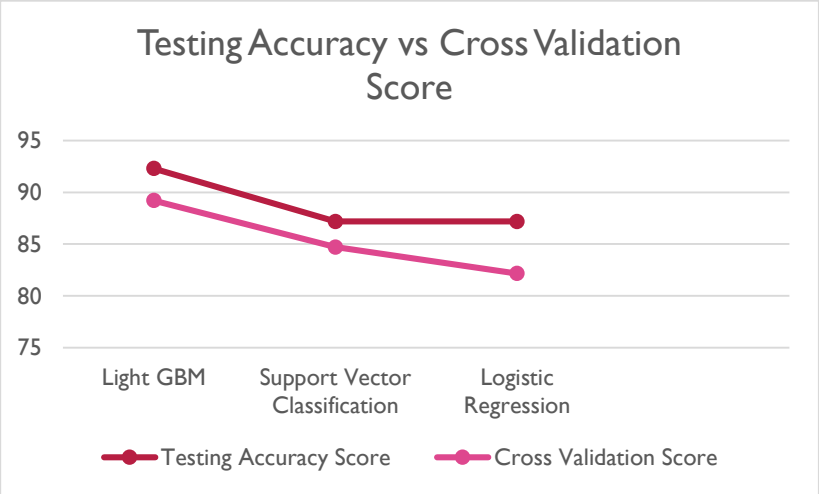


Figure 8.1: Testing Accuracy vs Cross Validation Score

This figure shows the comparison between the testing accuracy score and the cross-validation score of our three models. Where the test accuracy score in our predicted model is 92.31% and the cross-validation score is 89.21%.

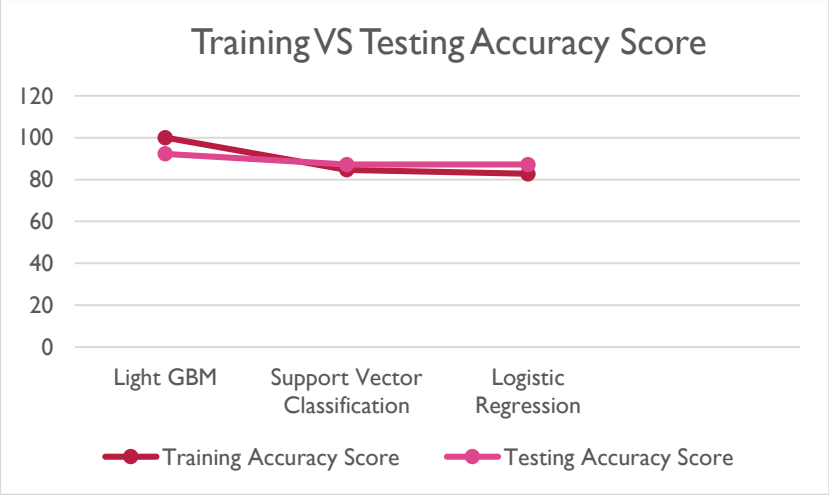


Figure 8.2: Training vs Testing Accuracy Score

This figure shows the comparison between the testing accuracy score and training accuracy score of our three models. Where the test accuracy score in our predicted model is 92.31% and the training accuracy score is 100%.

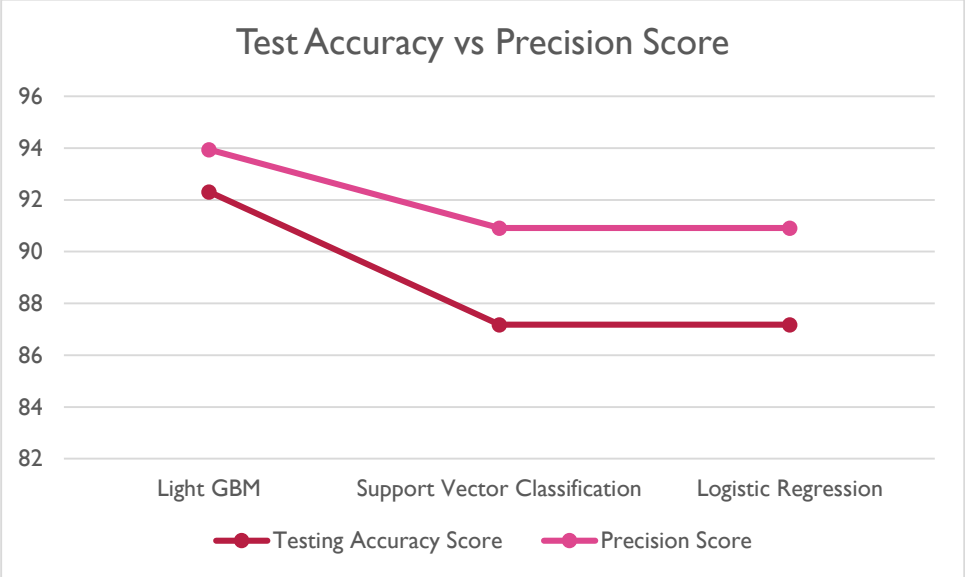


Figure 8.3 : Test Accuracy vs Precision Score

This figure shows the comparison between the testing accuracy score and precision score of our three models. Where the test accuracy score in our predicted model is 92.31% and the precision score is 93.94%.

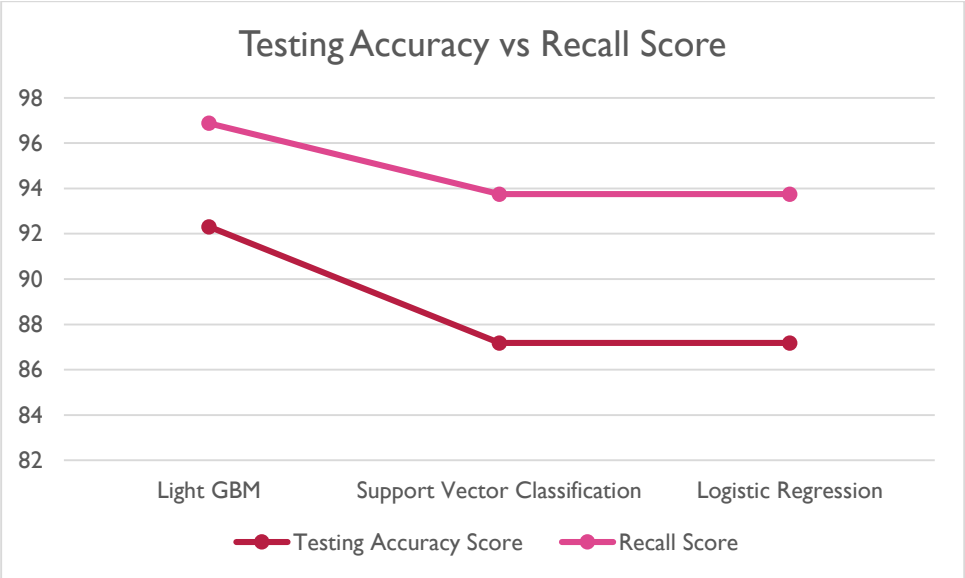


Figure 8.4 : Testing Accuracy vs Recall Score

This figure shows the comparison between the testing accuracy score and recall score for our three models. Where the test accuracy score in our predicted model is 92.31% and the recall score is 96.88%.

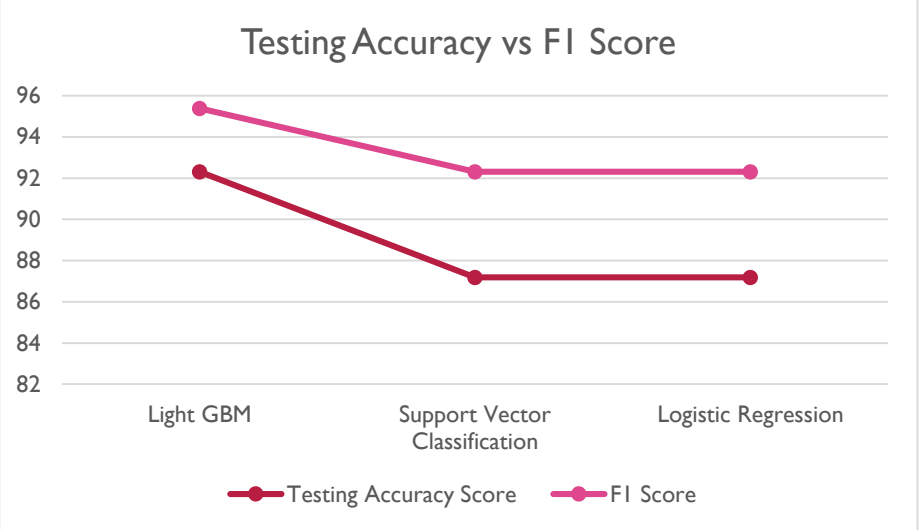
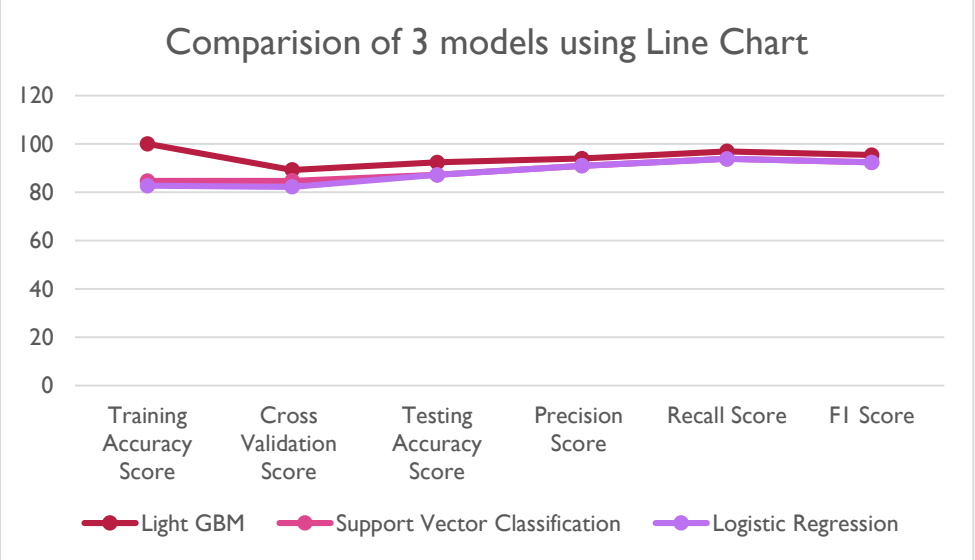


Figure 8.5: Testing Accuracy vs F1 Score



This figure shows the comparison between the testing accuracy score and f1 score for our three models. Where the test accuracy score in our predicted model is 92.31% and the f1 score is 95.38%.

Figure 8.6: Comparison of Three Models Using Line Chart

This figure shows the comparison of 3 models based on the five parameters i.e testing accuracy, cross validation, training accuracy, precision, recall and f1 scores. In our predicted model Light GBM secure with better testing accuracy score of 92.31%.

9. CONCLUSIONS:

In this research, we have addressed the significant issue of the freezing of gait (FOG) for patients having Parkinson's disease, a debilitating symptom that significantly impacts their quality of life. The ability to predict and treat freezing of gait early on is essential for enhancing patient outcomes because it is a complicated phenomenon with both motor and non-motors components. We proposed a predictive model for Freezing of gait using LightGBM, a powerful gradient boosting ensemble method known for its efficiency and accuracy. Using a large dataset of clinical characteristics, gait patterns, and demographic information of patients with Parkinson's disease, we employed feature engineering techniques to extract meaningful predictors associated with Freezing of gait. Our findings showed that the LightGBM model is capable of reliably predicting episodes of freezing of gait in Parkinson's patients. The model showed astounding precision, sensitivity, and specificity, suggesting that it could be a useful tool in clinical applications. Additionally, we carried out a thorough analysis of

comparable studies in the areas of anticipating freezing of gait and identifying Parkinson's disease. Highlighting various approaches, including machine learning, deep learning, and context recognition algorithms. These approaches contribute to the ongoing efforts to enhance our understanding of Parkinson's disease and improve patient care. The methodology section provided insights into the architecture and advantages of LightGBM, highlighting its effectiveness, lower memory use, and capacity for handling massive amounts of data. The leaf-wise tree growth method employed by LightGBM was explained, showcasing its advantages in terms of training speed and accuracy. In conclusion, our study adds to the corpus of work being done to address the problems caused by Parkinson's disorder and gait freezing. The predictive model developed in this study has the potential to assist healthcare professionals in early intervention and treatment planning for Parkinson's patients, ultimately leading to improved patient outcomes and quality of life. Future research in this field may further refine and expand upon the techniques and models discussed here, offering hope for better management of Parkinson's disease.

Declarations

List of Abbreviations:

- FOG: Freezing of Gait
- LightGBM: Light Gradient Boosting Machine
- PPMI: Parkinson's Progression Markers Initiative
- ST-GCN: Spatial-Temporal Graphical Convolution Network
- PD: Parkinson's Disease
- MV-SAE: Multi-Variant Stacked Auto Encoder
- SVM: Support Vector Machine
- KNN: k-Nearest Neighbours
- WM-STGCN: Weighted adjacency matrix with virtual connection and Multi-Scale Temporal Convolution in a Spatiotemporal Graph Convolution Network.
- LSTM: Long Short-Term Memory
- STGCN: Spatio Temporal Graph Convolutional Network
- ANN: Artificial Neural Networks
- RF-BFO-SVM: Relief Feature Selection Method Using Bacterial Foraging Optimization for Support Vector Machine
- PSO-SVM: Particle Swarm Optimization for Support Vector Machine
- KELM: Kernel Extreme Learning Machine
- RF: Random Forest
- SVC: Support Vector Classifier

Availability of data and material

The voice signals were captured and compiled into the dataset by Max Little of the University of Oxford.

Competing interests

The authors declare that they have no competing interests

Funding

No Funding

Authors' contributions

Single author has involved for the complete work implementation.

Acknowledgements

Not applicable

REFERENCES

[1] Esfahani, A. H., Dyka, Z., Ortmann, S., & Langendörfer, P. (2021). Impact of data preparation in freezing of gait detection using feature-less recurrent neural network. IEEE Access, 9, 138120-138131.

[2] Noor, M. H. M., Nazir, A., Ab Wahab, M. N., & Ling, J. O. Y. (2021). Detection of freezing of gait using unsupervised convolutional denoising autoencoder. IEEE Access, 9, 115700-115709.

[3] Lin, C. H., Wang, F. C., Kuo, T. Y., Huang, P. W., Chen, S. F., & Fu, L. C. (2022). Early detection of Parkinson's disease by neural network models. IEEE Access, 10, 19033-19044.

[4] Wang, W., Lee, J., Harrou, F., & Sun, Y. (2020). Early detection of Parkinson's disease using deep learning and machine learning. IEEE Access, 8, 147635-147646.

[5] Kwon, H., Clifford, G. D., Genias, I., Bernhard, D., Esper, C. D., Factor, S. A., & McKay, J. L. (2023). An explainable spatial-temporal graphical convolutional network to score freezing of gait in parkinsonian patients. Sensors, 23(4), 1766.

[6] Borzì, L., Sigcha, L., & Olmo, G. (2023). Context Recognition Algorithms for Energy-Efficient Freezing-of-Gait Detection in Parkinson's Disease. Sensors, 23(9), 4426.

[7] Nagasubramanian, G., Sankayya, M., Al-Turjman, F., & Tsaramirsis, G. (2020). Parkinson data analysis and prediction system using multi-variant stacked auto encoder. IEEE Access, 8, 127004-127013.

[8] Zhang, J., Lim, J., Kim, M. H., Hur, S., & Chung, T. M. (2023). WM–STGCN: A Novel Spatiotemporal Modeling Method for Parkinsonian Gait Recognition. Sensors, 23(10), 4980.

[9] Cai, Z., Gu, J., & Chen, H. L. (2017). A new hybrid intelligent framework for predicting Parkinson's disease. IEEE Access, 5, 17188-17200.

[10] Kour, N., & Arora, S. (2019). Computer-vision based diagnosis of Parkinson's disease via gait: A survey. IEEE Access, 7, 156620-156645..

..