PREDICTIVE MODELING OF PARKINSON'S DISEASE PROGRESSION USING LIGHTGBM CLASSIFIER

A project report submitted in partial fulfilment of the requirement for the award of degree of

BACHELOR OF TECHNOLOGY

In

INFORMATION TECHNOLOGY

Submitted by

Dola Shamitha	**20341A1221**
Kanithi Purna Chandu	**20341A1245**
Ganta Avinash	**20341A1224**
Duvvana Pavan Kumar	**20341A1222**

Under the esteemed guidance of

Mr. Ch. Anil Kumar

Asst. Professor, Dept of Information Technology

OCTOBER 2023

GMR Institute of Technology

An Autonomous Institute Affiliated to JNTU-GV, Vizianagaram

(Accredited by NBA, NAAC with 'A' Grade & ISO 9001:2015 Certified Institution)

GMR Nagar, Rajam – 532127,

Andhra Pradesh, India

2023 – 2024

GMR Institute of Technology
An Autonomous Institute Affiliated to JNTU-GV

GMRIT
Training Tomorrow's Engineers Today

Department of Information Technology

CERTIFICATE

This is to certify that the thesis entitled **PREDICTIVE MODELING OF PARKINSON'S DISEASE PROGRESSION USING LIGHTGBM CLASSIFIER**submitted by **DOLA SHAMITHA (20341A1221)**, **KANITHI PURNA CHANDU (20341A1245), GANTA AVINASH (20341A1224), DUVVANA PAVAN KUMAR (20341A1222)** has been carried out in partial fulfilment of the requirement for the award of degree of **Bachelor of Technology** in **Information Technology** of **GMRIT, Rajam** affiliated to **JNTU-GV, Vizianagaram** is a record of bonafide work carried out by them under my guidance & supervision. The results embodied in this report have not been submitted to any other University or Institute for the award of any degree.

Signature of Supervisor	**Signature of HOD**
Mr.Ch.Anil Kumar	Dr. V Vasudha Rani
Asst. Professor	Associate Professor & HOD
Department of IT	Department of IT
GMRIT, Rajam	GMRIT, Rajam

The report is submitted for the viva-voice examination held on ………………..

Signature of External Examiner

ACKNOWLEDGEMENT

It gives us an immense pleasure to express deep sense of gratitude to our guide **Mr. Ch. Anil Kumar,** Asst. Professor, Department of Information Technology of whole hearted and invaluable guidance throughout the project work. Without his sustained and sincere effort, this project work would not have taken this shape. He encouraged and helped us to overcome various difficulties that we have faced at various stages of our project work.

We would like to sincerely thank our Head of the Department **Dr. V Vasudha Rani**, for providing all the necessary facilities that led to the successful completion of our project work.

We would like to take this opportunity to thank our beloved Principal **Dr. C.L.V.R.S.V. Prasad**, for providing all the necessary facilities and a great support to us in completing the project work.

We would like to thank the project coordinator **Mr. Ch. Anil Kumar**, all the faculty members and the non-teaching staff of the Department of Information Technology for their direct or indirect support for helping us in completion of this project work.

Finally, we would like to thank all of our friends and family members for their continuous help and encouragement.

DOLA SHAMITHA	**20341A1221**
KANITHI PURNA CHANDU	**20341A1245**
GANTA AVINASH	**20341A1224**
DUVVANA PAVAN KUMAR	**20341A1222**

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 provides the best results with an accuracy of 92.31% when compared to other models like Support Vector Classifier and Logistic Regression.

Keywords: Parkinson's disease, Freezing of gait (FOG), LightGBM, Motor and non-motor symptoms, Gait patterns, Feature engineering, Gradient boosting.

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVATIONS

ADLs	:	ACTIVITIES OF DAILY LIVING
AE	:	AUTOENCODER
ANN	:	ARTIFICIAL NEURAL NETWORKS
API	:	APPLICATION PROGRAMMING INTERFACE
BMSAE	:	BALANCED MULTI-STACKED AUTO ENCODER
CFS	:	CORRELATION BASED FEATURE SELECTION
CLI	:	COMMAND-LINE INTERFACES
DFD	:	DATA FLOW DIAGRAM
DT	:	DECISION TREE
FN	:	FALSE NEGATIVE
FOG	:	FREEZING OF GAIT
FP	:	FALSE POSITIVE
GAE	:	GENES-AUTOENCODER
GNSS	:	GLOBAL NAVIGATION SATELLITE SYSTEM
GP-EM	:	GENETIC PROGRAMMING AND EXPECTATION MAXIMIZATION ALGORITHM
GPU	:	GRAPHICAL PROCESSING UNIT
GUI	:	GRAPHICAL USER INTERFACE
HCL	:	HARDWARE COMPATIBILITY LIST
HSSAE	:	HYPERACTIVE SUPERVISED STACKED AUTOENCODER
HUSAE	:	HYPERACTIVE UNSUPERVISED STACKED AUTOENCODER
IMU	:	INERTIAL MEASUREMENT UNIT
KNN	:	K-NEAREST NEIGHBORS
LGBM	:	LIGHT GRADIENT-BOOSTING MECHANISM
LR	:	LOGISTIC REGRESSION
LSTM	:	LONG SHORT-TERM MEMORY
MANN	:	MULTI-ATTRIBUTE ARTIFICIAL NEURAL NETWORK
MDS-UPDRS	:	MOVEMENT DISORDER SOCIETY UNIFIED PARKINSON'S DISEASE RATING SCALE
ML	:	MACHINE LEARNING
MRVMs	:	MULTI-CLASS MULTIKERNAL RELEVANCE VECTOR MACHINES
MSAEPD	:	MULTI STACKED AUTO ENCODER PARKINSON'S DISEASE
MVS-SAE	:	MULTI-VARIANT STACKED AUTO ENCODER
NN	:	NEURAL NETWORKS

OPF	:	OPTIMUM PATH FOREST CLASSIFIER
OS	:	OPERATING SYSTEM
PD	:	PARKINSON'S DISEASE
PPMI	:	PARKINSON'S PROGRESSION MARKERS INITIATIVE
PSO-SVM	:	PARTICLE SWARM OPTIMIZATION SVM
PwPD	:	PERSON WITH PARKINSON'S DISEASE
RF	:	RANDOM FOREST
RF-BFO-SVM	:	RELIEF FEATURE SELECTION USING BACTERIAL FORAGING OPTIMIZATION
SAE	:	STACKED AUTO ENCODER
SAS	:	SIMPSON-ANGUS SCALE
SDK	:	SOFTWARE DEVELOPMENT KIT
ST-GCN	:	SPATIAL-TEMPORAL GRAPHICAL CONVOLUTION NETWORK
SVC	:	SUPPORT VECTOR CLASSIFIER
TN	:	TRUE NEGATIVE
TP	:	TRUE POSITIVE
SVM	:	SUPPORT VECTOR MACHINE
UCI	:	UNIVERSITY OF CALIFORNIA
UI	:	USER INTERFACE
UMLBD	:	UNSUPERVISED ML FOR BIO VARIATION DETECTION
UPDRS	:	UNIFIED PARKINSON'S DISEASE RATING SCALE
WM-STGCN	:	WEIGHTED ADJACENCY MATRIX-STGCN

Chapter – 1

INTRODUCTION

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 trapped in 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 number has 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 mid line (such as the neck and trunk) or both sides of the body. Movement anomalies including tremors, rigidity, and others become more obvious. Those affected 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.

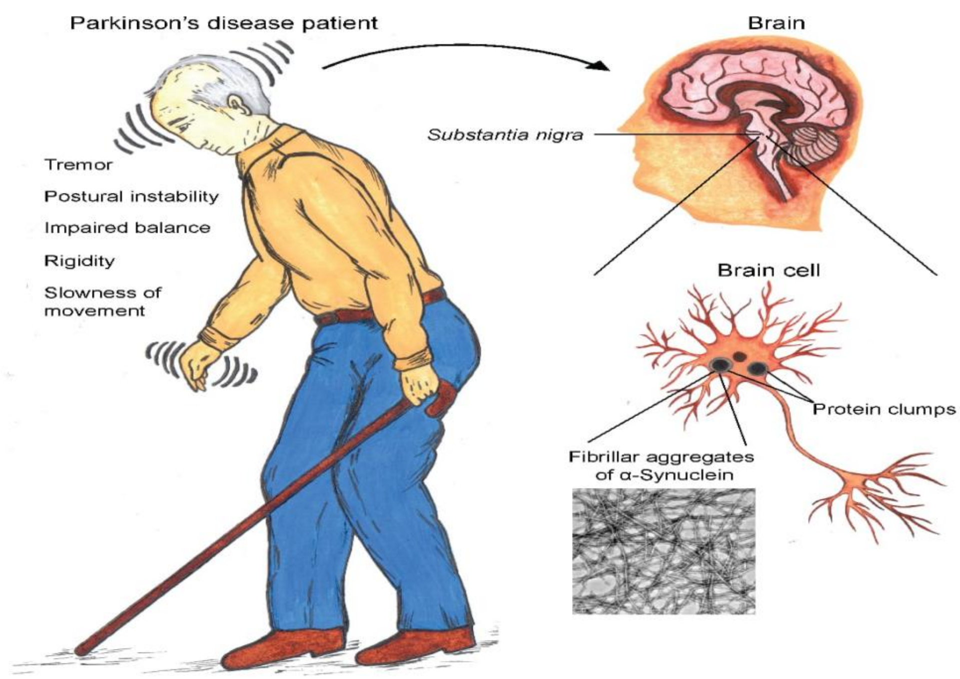


Figure 1.1 : Parkinson's Disease Symptoms

The continuing loss of brain cells and neurons in the substantial 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.

1.1 MATERIALS AND METHODS:

This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds to one of 195 voice recordings from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to the "status" column which is set to 0 for healthy and 1 for PD.

1.1.1 DATA COLLECTION:

The Parkinson's dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds to one of 195 voice recordings from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to the "status" column which is set to 0 for healthy and 1 for PD.

1.1.2 DISEASE PREDICTION:

Voice analysis is a promising new method for disease prediction, including Parkinson's disease (PD). It is based on the idea that the voice can be used to detect subtle changes in the body that may be indicative of disease. There are a number of different ways to analyze voice data for PD prediction. One common approach is to use machine learning to train a model to predict the presence of PD based on a set of voice features. For example, a model could be trained to predict the presence of PD based on features such as the pitch, loudness, and rhythm of the voice. Another approach to voice analysis for PD prediction is to look for specific biomarkers in the voice. For example, researchers have identified a number of biomarkers in the voice that are associated with PD. By tracking changes in these biomarkers over time, researchers may be able to identify people who are at risk of developing PD even before they experience any symptoms. Voice analysis is still a relatively new field, but it has the potential to revolutionize the way we diagnose and treat PD. By detecting PD early, voice analysis could help to improve patient outcomes and reduce healthcare costs.

Researchers at the University of California, San Francisco have developed a voice analysis test that can predict Parkinson's disease with 90% accuracy. The test is based on a machine learning algorithm that analyzes the pitch, loudness, and rhythm of the voice.

1.1.3 VOICE ANALYSIS:

Voice analysis is a promising new method for PD prediction. It is a non-invasive and relatively inexpensive method that could be used to screen large populations for PD. However, more research is needed to validate the accuracy and reliability of voice analysis tests.

Here are some of the benefits of using voice analysis for PD prediction:

1. It is a non-invasive and painless procedure.
2. It is relatively inexpensive.
3. It can be used to screen large populations for PD.
4. It can be used to detect PD early, even before symptoms appear.

However, there are also some limitations to using voice analysis for PD prediction:

1. The accuracy of voice analysis tests can vary depending on the specific test used and the population being tested.
2. Voice analysis tests may not be able to detect all cases of PD.
3. Voice analysis tests should not be used as the sole basis for diagnosing or treating PD.

1.1.4 MODEL SELECTION:

LightGBM is a gradient boosting algorithm that is often used for classification tasks. It is a fast and efficient algorithm that can handle large datasets. LightGBM has been shown to be effective in predicting Parkinson's disease (PD) using a variety of features, including demographic data, medical history, and clinical data. Model selection is an important step in developing a LightGBM model to predict PD. There are a number of factors to consider when selecting a model, such as the size and complexity of the dataset, the desired performance metrics, and the computational resources available. One common approach to model selection for LightGBM is to use a grid search. In a grid search, the model is trained and evaluated using a variety of different hyper parameter settings. The hyper parameter settings that produce the best performance on the evaluation set are then selected for the final model.

1.2 ENSEMBLE LERNING:

Ensemble learning is a machine learning technique that combines multiple machine learning models to produce a more accurate and robust model. Ensemble learning is based on the principle that a group of diverse models is likely to perform better than any individual model. There are a number of different ensemble learning algorithms. One common approach is to train a number of different machine learning models on the same training data. The predictions of these models are then combined to produce a final prediction. This can be done using a variety of different methods, such as averaging the predictions or voting on the predictions. Another common approach to ensemble learning is to use a bootstrapped aggregating (bagging) procedure. In bagging, the training data is resampled with replacement to create a number of new training sets. Each training set is then used to train a machine learning model. The predictions of these models are then combined to produce a final prediction. Ensemble learning has been shown to be effective in improving the performance of machine learning models on a variety of tasks, including classification and regression. Ensemble learning is particularly useful for tasks where the training data is limited or noisy.

1.2.1 LIGHTGBM CLASIFIER:

LightGBM has been shown to be effective in ensemble learning, where it is combined with other machine learning models to produce a more accurate and robust model. One common approach to using LightGBM in ensemble learning is to train a number of different LightGBM models on the same training data, each with different hyper parameters. The predictions of these models are then combined to produce a final prediction. This can be done using a variety of different methods, such as averaging the predictions or voting on the predictions. Another common approach to using LightGBM in ensemble learning is to use a stacking procedure. In stacking, the predictions of a number of different machine learning models are used to train a new machine learning model. This new model is then used to make predictions on new data. LightGBM is a well-suited algorithm for ensemble learning because it is fast and efficient. It can also be used to train a variety of different models with different hyper parameters. This makes it easy to create diverse ensembles that are likely to perform well on new data.

1.2.2 WORK FLOW OF LIGHTGBM MODEL:

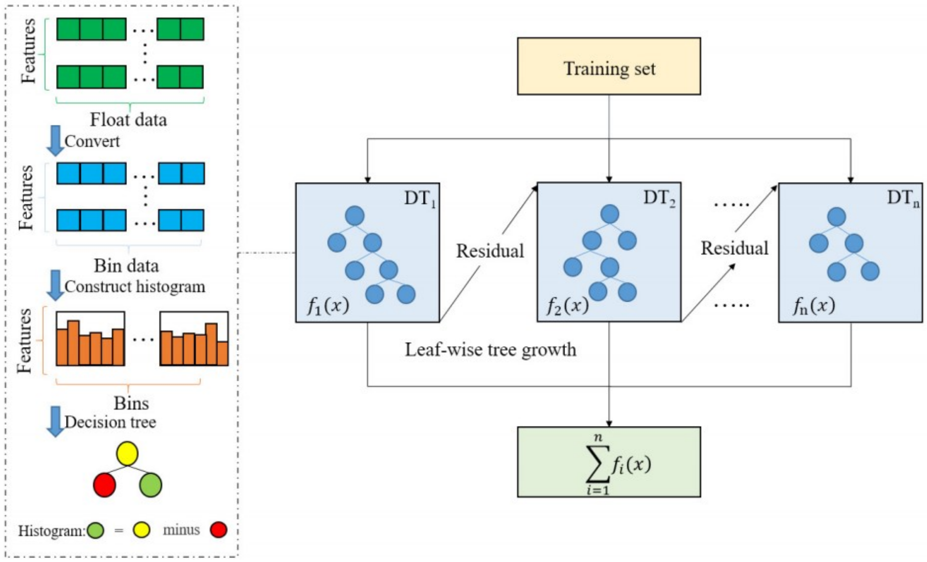


Figure 1.2 : LightGBM Workflow

The output data were first used to initialize a regression decision tree (DT) based on the loss function. Following that, *n* regression decision trees were created. A negative gradient, or pseudo-residual, was calculated using the output value of the initialized DT with the set sample output value. This residual was then used as the output value of the set sample to train the first DT. The best split point needed to be found when training the DT, and LightGBM discretized the continuous floating-point feature values and generated a histogram (bins). As the data were traversed, the cumulative statistics of each discrete value in the histogram were counted. Feature selection was accomplished by traversing the histogram based on the discrete values to determine the best split point. DT growth employs a leaf-wise technique, which locates and divides the leaf node with the maximum splitting gain relative to the present leaf node. In addition, during DT construction, the histogram of a leaf node can be obtained from the difference between its parent node's histogram and its sibling node's histogram. When DT reaches a depth limit or a limit on the number of leaf nodes, the residuals are fitted by assigning leaf node weights to fit this time. After training, the LightGBM model for prediction in the event of a GNSS outage can be updated, and the process can be repeated until *n* regression decision trees are constructed.

1.2.3 ADVANTAGES OF LIGHTGBM:

LightGBM has a number of advantages over other gradient boosting algorithms, including:

1. Speed: LightGBM is one of the fastest gradient boosting algorithms available. This is due to a number of factors, including its efficient data structures and algorithms.
2. Memory efficiency: LightGBM is also very memory efficient. This makes it suitable for training and deploying models on large datasets.
3. Accuracy: LightGBM has been shown to be as accurate as, or even more accurate than, other gradient boosting algorithms on a variety of tasks.
4. Scalability: LightGBM can be scaled to handle very large datasets. It can also be distributed across multiple machines to further improve performance.

In addition to these advantages, LightGBM is also relatively easy to use. It has a simple and intuitive API, and there are a number of tutorials and resources available online. Overall, LightGBM is a powerful and versatile machine learning algorithm that can be used for a variety of tasks. It is a good choice for anyone looking for a fast, efficient, and accurate machine learning algorithm.

1.2.4 LIMITATIONS OF LIGHTGBM:

LightGBM is a powerful machine learning algorithm with a number of advantages, but it also has some limitations. Here are a few to be aware of:

1. Over-fitting: Like any machine learning algorithm, LightGBM is susceptible to over-fitting. This means that the model can learn the training data too well and may not generalize well to new data. To prevent over-fitting, it is important to use regularization techniques and to validate the model on a held-out test set.
2. Interpretability: LightGBM models can be difficult to interpret, especially for complex models with many trees. This makes it difficult to understand why the model makes the predictions that it does.
3. Hyperparameter tuning: LightGBM has a number of hyper-parameters that can be tuned to improve the model's performance. However, hyper-parameter tuning can be time-consuming and difficult, especially for large and complex models.
4. Computational resources: LightGBM can be computationally expensive to train on large datasets. This may require access to powerful hardware and software.

1.3 LOGISTIC REGRESSION:

Logistic regression is a statistical model that can be used to predict the probability of a binary outcome. For example, logistic regression can be used to predict the probability of a customer churning, whether or not a patient has a disease, or whether or not a product will be a success. Logistic regression works by fitting a logistic function to the data. The logistic function is a sigmoid function that maps input values to output values between 0 and 1. The output value of the logistic function represents the probability of the binary outcome. Logistic regression is a simple and interpretable model, which makes it a popular choice for machine learning practitioners. It is also a relatively efficient algorithm to train and deploy.

1.4 SUPPORT VECTOR CLASSIFIER:

A support vector classifier (SVC) is a type of machine learning algorithm used for classification. It works by finding a hyperplane in the data that separates the different classes with the largest possible margin. The hyperplane is a line or plane in the data space that divides the data into two or more categories. The margin is the distance between the hyperplane and the closest data points from each class. SVC is trained on a labeled dataset, where each data point has a known class label. The algorithm learns to find the hyperplane that best separates the data points into their respective classes. Once the model is trained, it can be used to classify new data points by assigning them to the class of the closest data points in the training dataset. SVC is a powerful and versatile machine learning algorithm that can be used for a variety of classification tasks. They are particularly well-suited for problems with high-dimensional data and non-linear relationships between the features.

1.5 MODEL PREDICTIONS:

Model predictions for Parkinson's disease (PD) prediction can vary depending on the model used, the quality of the data, and the stage of the disease. However, some studies have shown that machine learning models can achieve high accuracy in predicting PD, even at early stages. LightGBM model has been used and the evolution metrics used are Accuracy, Precision, Recall, F1-score and Cross Validation.

Chapter – 2

LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 Study of Related Papers

Following research papers are studied in detail to understands the proposed recommendation technique and experimental result for predicting the output.

Here we have gathered several periodicals that have conducted research on our connected work, which is based on Parkinson's disease, and we have separately summarized each work as shown below.

Paper 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.

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.

This work is based on the stateless definition of LSTM, i.e., the model passes states to the next time step in the same window and it does not transmit any states to the further windows. The input data are fed to the LSTM model with 110 hidden units in the first layer which is followed by two additional layers with 90 and 70 hidden units. The whole model consists of 167600 learnable parameters. The type of the model is many-to-one, so the input data is a three-channel raw data from the acceleration sensor mounted on the shank and the output is a label defining the class of the input data. The LSTM layers are followed by a fully connected layer, and then the Soft-Max layer assigns the likelihoods of the outputs to one of the two classes. Finally, the window is classified as FOG or Normal class.

Paper 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 auto-encoder. IEEE Access, 9, 115700-115709.

The unsupervised convolution denoising auto-encoder model works by first per-processing the gait data from wearable sensors to remove noise. Then, the model learns to extract features from the per-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.

This paper proposes a feature learning method using deep denoising autoencoder for detecting FOG episodes in PD gait acceleration data. The proposed method automatically learns feature representation of the data in an unsupervised manner, eliminating the need for manually handcrafted feature engineering. The deep denoising autoencoder is trained to minimize the cost function with Kullback-Leibler divergence, which improves the reconstruction outputs. As a result, more salient features can be learned, improving the accuracy of the FOG detection model. We evaluate the proposed method on a benchmark Daphnet dataset. The results showed that the proposed method managed to achieve 90.94% sensitivity and 67.04% specificity. These results are comparable to theoriginal Daphnet dataset research. The low specificity results are due to the significant lack of data from the FOG class label compared with non-FOG class label data, which hinders feature learning of the deep denoising autoencoder. The proposed autoencoder exploits the capability of convolution and pooling layers to leverage temporal structure and learn the salient features of gait data. Furthermore, the deep denoising autoencoder can produce a compact feature representation, therefore reducing the over-fitting of the data. The proposed method is evaluated using a benchmark public dataset. The experimental results show that the proposed method can achieve a high detection accuracy and its performance is comparable, if not better, than the existing methods.

Paper 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.

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 convolution 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.

This paper develops neural network models that can recognize Parkinson's disease (PD) at its early stage. PD is a common neurodegenerative disorder that presents with progressive slow movement, tremor, limb rigidity, and gait alterations, including stooped posture, shuffling steps, festination, freezing of gait, and falling. Early detection of PD enables timely initiation of therapeutic management that decreases morbidity. However, correct recognition of PD, especially in early-stage disease, is challenging because the aging population, which has a high PD prevalence, also commonly exhibits progressive gait slowness due to other disorders, such as joint osteoarthritis or sarcopenia. Therefore, developing a reliable and objective method is crucial for differentiating PD gait characteristics from those of the normal elderly. The aim of this study was to develop neural network models that could use the participants' motion data during walking to identify PD. We recruited 32 drug-naive PD patients with variable disease severity and 16 age/sex-matched healthy controls, and we measured their motions using inertial measurement unit (IMU) sensors. The IMU data were used to develop neural network models that could identify patients with advanced-stage PD with an average accuracy of 92.72% in validation processes. The models also differentiated patients with early-stage PD from normal elderly subjects with an accuracy of 99.67%. Another independent group of participants recruited to test the developed models confirmed the successful discrimination of PD-affected from healthy elderly, as well as patients at different severity stages. Our results provide support for early diagnosis and disease severity monitoring in patients with PD. The developed model structures might also be applied to the classification of other diseases, such as the discrimination between benign tremors and early PD, if enough samples are available.

Paper 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.

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%.

Accurate and early detection of PD is vital due to its ability to provide crucial information to slow down the progression of PD. All over the years, various data-driven methods have been developed to improve the detection of PD. In contrast to the model-based detection techniques, where prior availability of an analytical model is required, in data-driven techniques, only the availability of historical data is needed. Recently, machine learning (ML) has emerged as a promising field of research in PD diagnosis, both in academia and industry . Owing to its data-driven approaches, ML has brought a paradigm shift in the way relevant information in PD biomarkers are extracted and analyzed. Furthermore, machine learning techniques provide pertinent information that offers guidance related to PD classification and diagnosis to speed up decision making. Various machine learning techniques have been applied in the literature to address the PD detection problem. For instance, in dysphonia measurements have been used to detect patients with PD from healthy people. The support vector machine (SVM) is applied to only four dysphonic features for PD classification due to its ability to extract non-linearity by using nonlinear kernels. In, three common machine learning algorithms, namely Random Forest (RF) or Support Vector Machine (SVM) and neural network, have been applied to detect Parkinson's disease based on acoustic analysis of speech. It has been shown the promising results of RF an SVM in early PD detection. In , the performance of four classifiers, Decision Trees, Regression and Neural Networks (NN), has been compared in detecting PD, and the best accuracy of 92.9% is obtained using NN algorithm. It has been shown that LSTM outperforms the SVM in detecting FOG.

Paper 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.

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.

A major innovation of our study is that it is the first study of its kind that uses the largest sample size (>30 h, N = 57) in order to apply explainable, multi-task deep learning models for quantifying FOG over the course of the medication cycle and at varying levels of parkinsonism severity. We trained interpretable deep learning models with multitask learning to simultaneously score FOG (cross-validated F1 score 97.6%), identify medication state (OFF vs. ON levodopa; cross-validated F1 score 96.8%), and measure total PD severity (MDSUPDRS-III score prediction error ≤ 2.7 points) using kinematic data of a well-characterized sample of N = 57 patients during levodopa challenge tests. The proposed model was able to explain how kinematic movements are associated with each FOG severity level that were highly consistent One critical factor limiting our ability to treat FOG is that clinicians measure it relatively coarsely, primarily with expert rater observations as part of the Movement Disorder Society Unified Parkinson's Disease Rating Scale Part III (MDS-UPDRS-III) scale . This scale requires specially trained raters who have typically completed movement disorders training. In addition, despite being resource-intensive, FOG is only quantified with a single item on an ordinal scale from 0 to 4, which may be too insensitive to detect small beneficial effects. The most established self-reporting scale used in research settings, the N-FOG-Q is acknowledged to be insufficiently sensitive for clinical trial use. Previous work have shown that FOG may be associated with non-dopaminergic system changes which suggests the potential for new treatments beyond dopaminergic medications like carbidopa-levodopa . However, developing a novel drug that is effective in treating FOG requires accurately quantifying FOG to increase the precision for clinical trials.

Paper 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.

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.

Motion sensors (e.g., accelerometer and gyroscopes) are small and lightweight and hence can be worn by the patient in daily life. They are inexpensive and allow accurate assessment of human movement in the laboratory, at home, and outside of the house . In recent decades, wearable inertial sensors have been used for a large number of medical applications , including FOG monitoring. In this context, many solutions have been proposed based on one or more sensors positioned on different parts of the body to record motion data and automatically detect FOG . However, the accuracy of FoG detection strongly depends on the efficacy of the addressed data processing approaches . For this reason, increasing attention has been devoted over the past decades to the development of effective algorithmic approaches for the automatic identification of FOG .

This study evaluates the effects of contextual algorithms applied to FOG detection. Specifically, four algorithmic approaches with different levels of complexity were designed for the detection of gait and activity. Performance was evaluated on a dataset comprising fifty PwPD performing gait and ADLs. The impact of context awareness on FOG detection performance was assessed in two different datasets comprising thirty-one PwPD performing a large number of walking tasks and ADLs and including more than 1200 FOG episodes. In this work, data recorded by a single accelerometer placed on the lower back were analyzed. This represents a simple and unobtrusive sensor configuration for passive long-term monitoring of PD. The results indicate that the use of a single inertial sensor and the implementation of context-aware approaches appear to be a viable option for implementing ecological and energy-efficient solutions for long-term FOG monitoring in ambulatory and free-living settings.The results indicate that a context classifier can reduce the computational burden of FOG detection algorithms without significantly affecting the FOG detection rate.

Paper 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.

This paper proposes a novel method for analyzing and predicting Parkinson's disease (PD) data. The proposed method, called MVS-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 MVS-SAE on the UCI Parkinson's Disease Data Set, which consists of 20 features from 195 PD patients and 195 healthy controls. MVS-SAE achieved an accuracy of 90.8%, which outperformed the baseline methods, including SVM, Random Forest, and KNN. The authors conclude that MVS-SAE is an effective method for analyzing 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 MVS-SAE could be used for clinical applications, such as early diagnosis and treatment of PD.

The Proposed MSAEPD System is implemented with various types of Auto Encodes (AEs) such as Classless SAE, Clustered SAE, Multi-level Balanced SAE and Multi-Variant SAE techniques. The Classless SAE, Clustered SAE are used for making trained data features under unsupervised and supervised categories respectively. Multi-level Balanced SAE and Multi-Variant SAE are used for feature classification based on trained Parkinson data. Multi-Variant is the term used to indicate the multiple PD features (Voice, motor and non-motor). These are varying for each iteration for each patient. Multi-Variant SAE is constructed with multiple parallel AEs to handle these PD features

The Proposed MSAEPD system was developed to ensure more accurate PD detection and PD optimal treatment assistantship. To achieve accurate PD classification, MSAEPD system used four different strategies with the help of complicated MSAE. In this work, HSSAE, HUSAE, BMSAE an VSAE were proposed and examined using real-time Parkinson dataset. The Proposed procedures are compared with MANN, GAE and UMLBD techniques using various critical parameters. In this comparison, MSAEPD system showed notable improvement than existing systems in all aspects. In this regard, the Proposed system is developed using Multi-Variant Stacked Auto Encoder (MVSAE). The MVSAE based PD Prediction System (MSAEPD) helps to analyze more PD symptoms than existing systems. This article provides four different variants of SAE construction procedures to predict PD symptoms. The MSAEPD is implemented and compared with existing works such as MANN, GAE and UMLBD.

Paper 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.

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%.

Clinical gait assessment is a commonly used method for performing gait analysis, which is an assessment performed by a clinician. Specifically, the physician needs to observe the patient’s walking performance and then give a score based on criteria of the Unified Parkinson’s Disease Rating Scale (UPDRS) and Simpson–Angus Scale (SAS) . Moreover, utilizing different types of sensors is a popular method. For example, sensors are embedded in the shoe insoles to measure the pressure of the foot against the ground while walking, inertial measurement units and goniometers are fixed to joints, such as the waist and elbow, to measure the walking speed and acceleration . Moreover, some studies have proposed video-based methods . For example, reflective markers are attached to diverse locations on the human body. The location and trajectory of the markers are analyzed to provide kinematic information by recording with a digital camera. The Vicon Vantage system requires about 8–14 high-precision cameras to provide accurate 3D motion data for gait analysis

In this paper, we proposed a novel spatiotemporal modeling approach, known as WM–STGCN, which employs a weighted adjacent matrix with virtual connections and multi-scale temporal convolutional networks to recognize Parkinsonian gait from forward walking videos. Our experimental results demonstrated the effectiveness of the proposed method, which outperformed the machine learning-based methods such as LSTM, KNN, Decision Tree, AdaBoost, and ST–GCN. This method could provide a promising solution for PD gait recognition, which is crucial for the early and accurate diagnosis of PD. We believe that our method can be further improved by integrating it with other advanced deep learning techniques.

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

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.

The results have shown that the ANN yielded the most accurate diagnostic results with an overall accuracy of 92.9%. In addition, Sakar and Kursun developed a PD diagnostic tool with a classification accuracy of 92.75% using a combination of mutual information and SVM. In another study, Psorakis proposed an improved multi-class multikernel relevance vector machines (mRVMs) to detect the PD. It yielded an overall PD classification accuracy rate of 89.47%. Guo et al. [9] proposed to use genetic programming and the expectation maximization algorithm (GP-EM) to develop a diagnostic tool for PD with a classification accuracy of 93.1%. More recently, Luukka employed the similarity classifier combined with fuzzy entropy measurements-based feature selection to detect PD; the resulting diagnostic method achieved a mean classification accuracy of 85.03% using only two dysphonic features. They developed a PD diagnostic tool with a classification accuracy of 93.47% using a fuzzy-based non-linear transformation method and SVM classifier. In another study, Ozcift and Gulten proposed to use the rotation forest ensemble classifier combined with a correlation based Feature selection (CFS) algorithm to identify patients with PD; the resulting model has produced a classification accuracy of 87.13%. In addition, AStrom and Koker developed a parallel feed-forward neural network with a PD classification accuracy of 91.20%. Also employed evolutionary-based techniques and the Optimum Path Forest (OPF) classifier to identify PD patients. In addition, the computational efficiency could be improved by parallel computing methods, and the proposed method could be applied to other medical diagnostic problems in future studies.

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

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.

The most common degenerative neurological disorder occupying the second place next to Alzheimer's disease i.e. Parkinson disease came into light after being described by a physician, James Parkinson in 1817. PD is a chronic disorder that progresses gently with time and has more life risk on the male population than females . PD being a syndrome of the nervous system directly affects the functioning of the brain and results in loss of neuromuscular control . An area in the human brain named as substantia nigra consists of dopaminergic neurons (transmitter neurons) that releases dopamine chemical to basal ganglia (receptor cell)

The obtained literature revealed the prime focus on VB markerless technology (about 48%) where the Kinect sensor has been used the most for PD analysis due to its depth detail catching capabilities. The article precisely surveyed the Preprocessing methods used to prepare PD gait data and also explored different categories of gait features that can be beneficial for PD gait evaluation. Among all, the fusion of PD gait features has been dominantly utilized as it enhances accuracy and provides a broader view for PD inspection

Several PD gait feature extraction and selection approaches are discussed. Data indicated the majority of search towards the use of PCA (almost 24%) for dimensional reduction. Also, the article surveyed Machine learning techniques that have been used and SVM classifier is analyzed to be most adopted by researchers to classify PD and normal subjects (approx. 23%) to provide more effective decision-making This article performed a systematic search of the existing literature to gather the data from reputed journals and conferences. Applying a number of search keywords yielded about 1500 related articles out of which 71 relevant articles that mainly focused on gait based PD detection using vision-based technology are selected after illuminating the irrelevant and duplicate ones

Chapter – 3

REQUIREMENT SPECIFICATION

3 . REQUIREMENT SPECIFICATION

3.1 Software Requirements

The software requirements are description of features and functionalities of the target system. Requirements convey the expectations of users from the software product. The requirements can be obvious or hidden, known or unknown, expected or unexpected from client's point of view.

3.1.1 Operating System

Windows is a graphical operating system developed by Microsoft. It allows users to view and store files, run the software, play games, watch videos, and provides a way to connect to the internet. It was released for both home computing and professional works.

MacOS is the computer operating system (OS) for Apple desktops and laptops. It is a proprietary graphical OS that powers every Mac. OSes interact with a computer's hardware, allocating the resources necessary to complete tasks given to it, for example, running an application. OSes allocate resources including memory, processing power and file storage.

Linux is an operating system. In fact, one of the most popular platforms on the planet, Android, is powered by the Linux operating system. An operating system is software that manages all of the hardware resources associated with your desktop or laptop. To put it simply, the operating system manages the communication between your software and your hardware. Without the operating system (OS), the software wouldn't function.

3.1.2 SDK

SDK stands for software development kit or devkit for short. It's a set of software tools and programs used by developers to create applications for specific platforms. SDK tools will include a range of things, including libraries, documentation, code samples, processes, and guides those developers can use and integrate into their own apps. SDKs are designed to be used for specific platforms or programming languages.

Some of SDK's used in this project are stated below:

1. TensorFlow

2. Streamlit
3. Numpy
4. Scikit-Learn etc.

3.2 Hardware Requirements :

The most common set of requirements defined by any operating system or software application is the physical computer resources, also known as hardware, A hardware requirements list is often accompanied by a hardware compatibility list (HCL), especially in case of operating systems. An HCL lists tested, compatible, and sometimes incompatible hardware devices for a particular operating system or application. The following sub-sections discuss the various aspects of hardware requirements. The Hardware Interfaces Required are:

1. Ram: Minimum 8GB or higher
2. GPU: 4GB dedicated
3. SSD: 128GB
4. Processor : Intel i5 10th Gen or Ryzen 5 with Octa core.

3.3 Non-Functional Requirements:

Non-Functional Requirements are the constraints or the requirements imposed on the system. They specify the quality attribute of the software. Non- Functional Requirements deal with issues like scalability, maintainability, performance, portability, security, reliability, and many more. Non-Functional Requirements address vital issues of quality for software system. It includes below things: Capacity, Availability andPerformance etc.

3.4 Python Libraries To be installed:

The following libraries with the specific versions type must be installed for this project to function. These can be installed using command “pip install library_name==version"

- numpy==1.23.3
- Pillow==9.2.0

- scikit-learn==1.1.2
- scipy==1.9.2
- sklearn==0.0
- tensorboard==2.10.0
- tensorflow==2.10.0
- tensorflow-estimator==2.10.0
- tensorflow-gpu==2.10.0
- tensorflow-hub==0.12.0

Chapter – 4

METHODOLOGY

4. METHODOLOGY

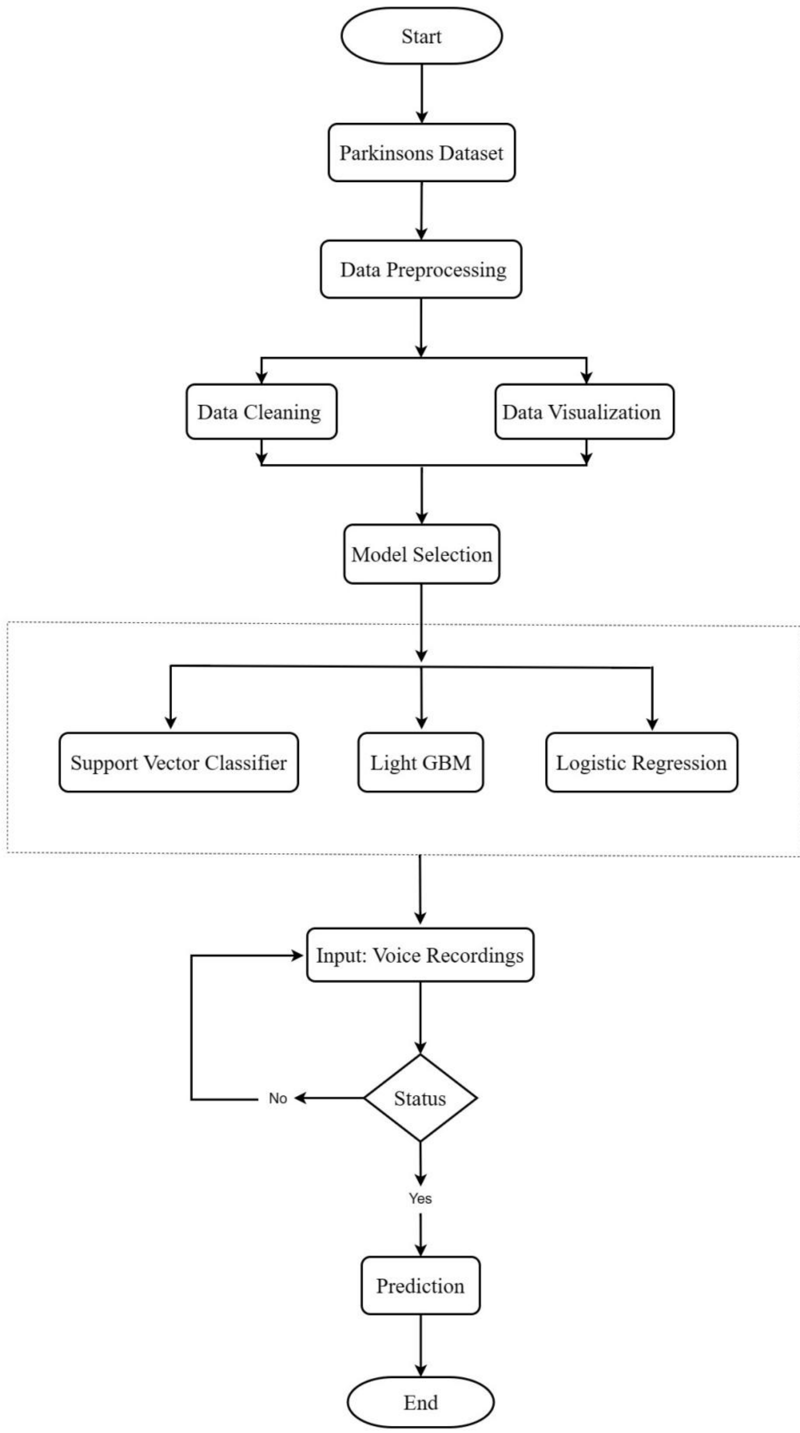


Figure 4.1 : Model Workflow

4.1 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.

4.2 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 4.2 : This dataset consists of 195 records and 24 attributes based on the patients vocal frequency.

4.3 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 the algorithms.

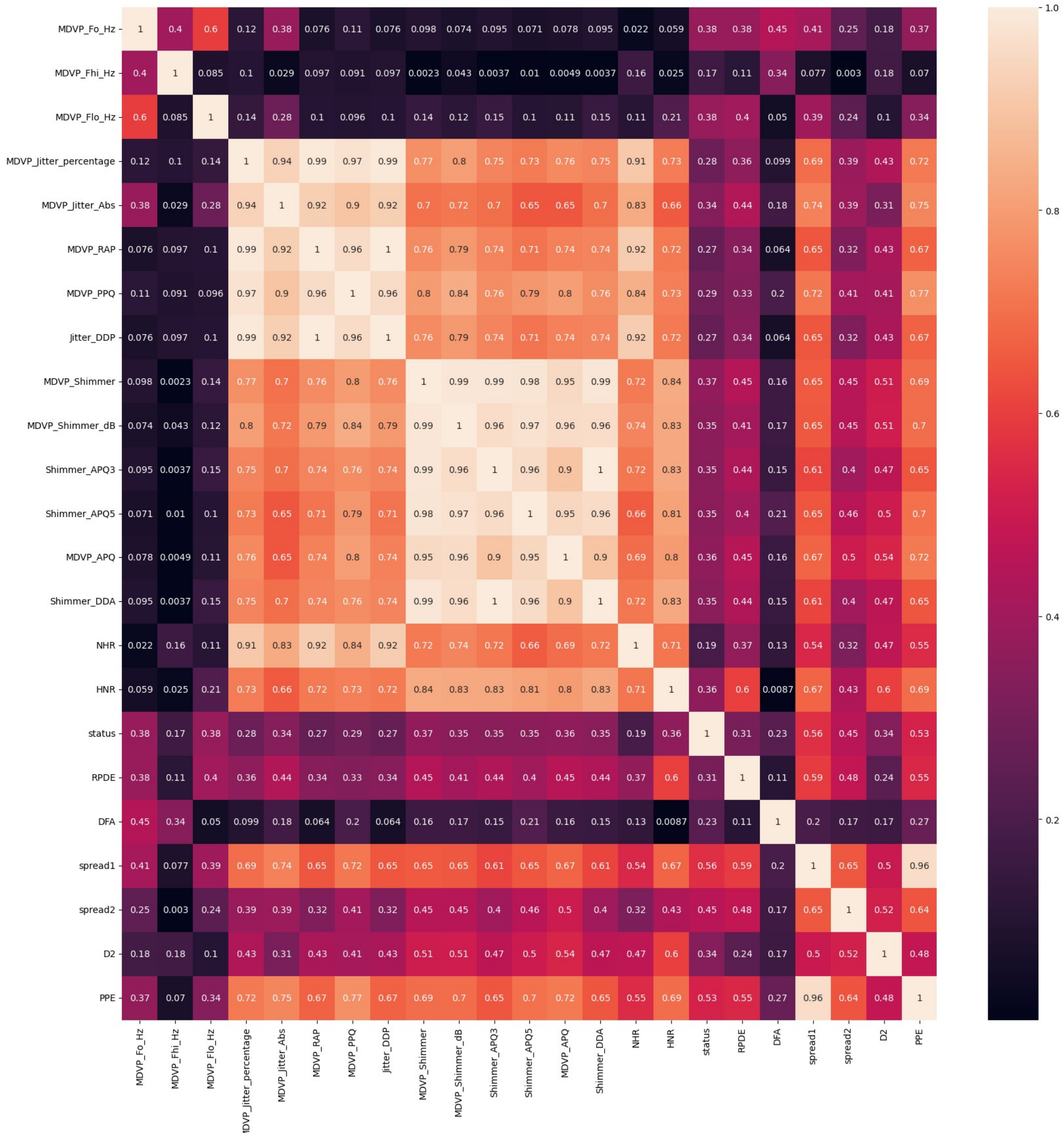


Figure 4.3 : Correlation between the attributes

4.4 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.

4.5 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.

4.6 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.

4.7 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.

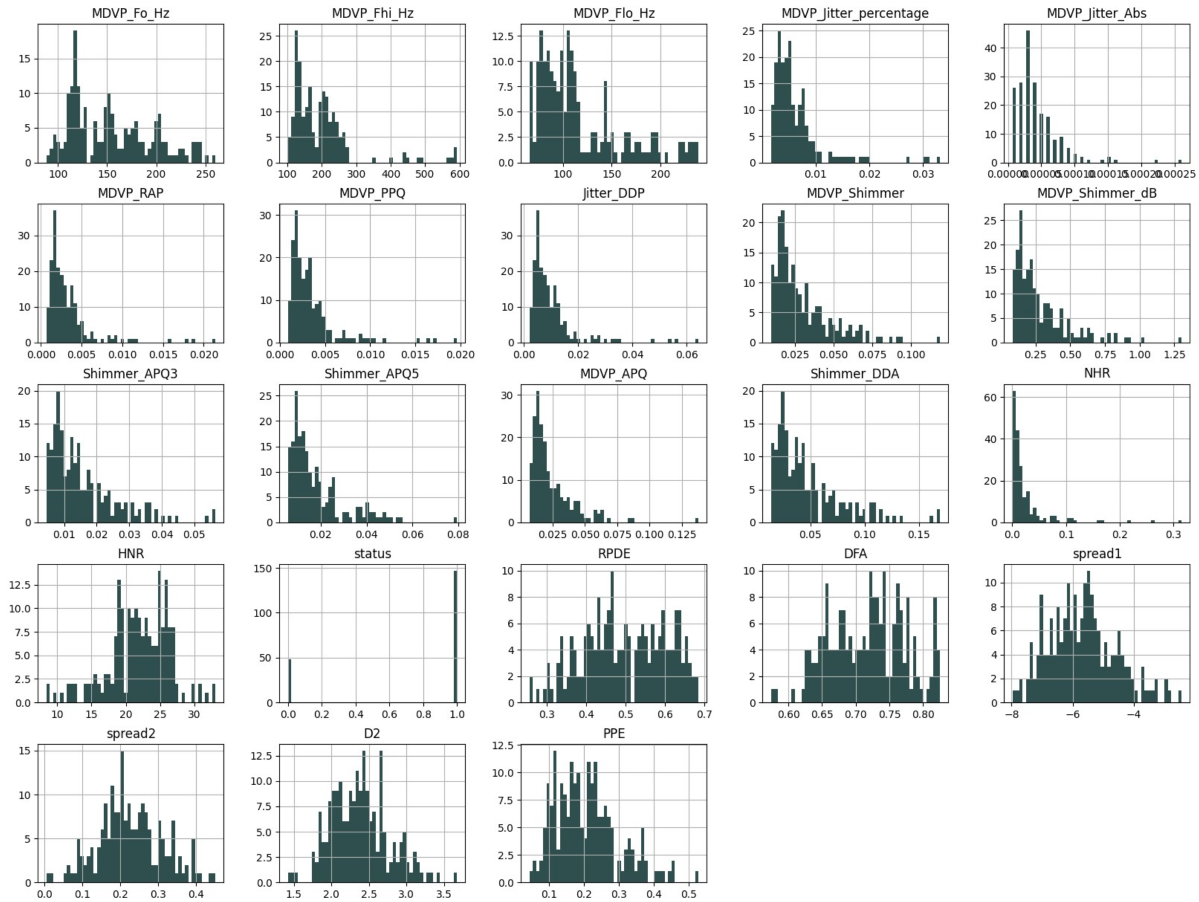


Figure 4.4 : Representation of the frequencies of each attribute

4.8 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.

4.8.1 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.

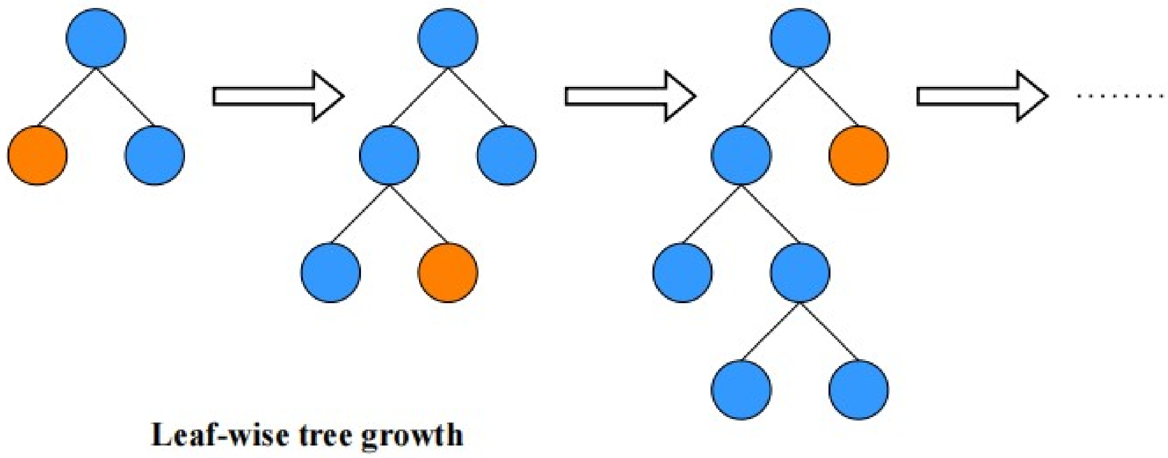


Figure 4.5 : Architecture of LGBM

Best-first (leaf-wise) and depth-first (level-wise) will produce the same tree if we develop theentire tree. Order is important because trees aren't typically grown to their full depth. 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 convergeto the identical performance without pausing or trimming.

4.8.2 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 SVM.

4.8.3 Logistic Regression:

The logistic regression is used 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

4.9 Model Evalution:

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.

4.9.1 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.

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

4.9.2 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.

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

4.9.3 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.

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

4.9.4 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 = $\frac{2.Precision.Recall}{Precision+Recall}$

4.9.5 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.

4.10 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 to use in converting data scripts into shareable web apps. 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.

4.10.1 Creating Streamlit Framework:

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 4.6 : Installing Streamlit

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

```
Requirement already satisfied: mdurl~=0.1 in c:\users\jagadeesh\appdata\local\packages\pythonsoftwar
efoundation.python.3.10_qbz5n2kfra8p0\localcache\local-packages\python310\site-packages (from markdo
wn-it-py>=2.2.0->rich<14,>=10.14.0->streamlit) (0.1.2)
PS C:\Users\Jagadeesh\OneDrive\Desktop\Project\Parkinson-disease-detection-main\Parkinson-disease-de
tection-main>
PS C:\Users\Jagadeesh\OneDrive\Desktop\Project\Parkinson-disease-detection-main\Parkinson-disease-de
tection-main> python -m streamlit run parkinsons_detection.py

  You can now view your Streamlit app in your browser.

  Local URL: http://localhost:8501
  Network URL: http://192.168.137.56:8501
```

Figure 4.7 : Running Streamlit

4.11 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.

4.11.1 User Interface:

A user interface (UI) is the graphical layout of an application that allows users to interact with it. It includes all the elements that users see and interact with, such as buttons, menus, text boxes, and images. The UI is designed to be easy to use and navigate, and to provide users with a positive experience when interacting with the application. There are two main types of UIs: graphical user interfaces (GUIs) and command-line interfaces (CLIs). GUIs use visual elements to represent the application and its data, while CLIs use text-based commands. GUIs are the most common type of UI, and are used in most modern applications. They are typically more user-friendly than CLIs, as they do not require users to learn complex commands. However, CLIs can be more efficient for certain tasks, such as system administration.

Chapter – 5

SYSTEM DESIGN

5. SYSTEM DESIGN

5.1 System Overview

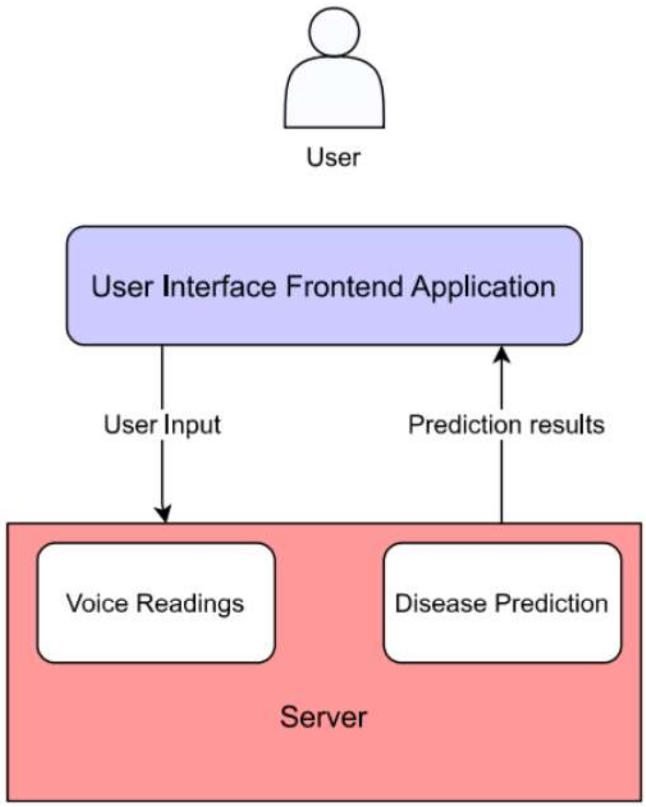


Figure 5.1 : Overview of System Architecture

5.2 Data Flow Diagram:

The DFD is also known as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of the input data to the system, various processing carried out on these data, and the output data is generated by the system. It maps out the flow of information for any process or system, how data is processed in terms of inputs and outputs. It uses defined symbols like rectangles, circles and arrows to show data inputs, outputs, storage points and the routes between each destination. They can be used to analyses an existing system or model of a new one. A DFD can often visually "say" things that would be hard to explain in words and they work for both technical and non- technical. There are four components in DFD

1. External Entity
2. Process
3. Data Flow
4. Data Store

External Entity:

It is an outside system that sends or receives data, communicating with the system. They are the sources and destinations of information entering and leaving the system. They might be an outside organization or person, a computer system or a business system. They are known as terminators, sources and sinks or actors. They can be used to analyses an existing system or model of a new one. These are sources and destinations of the system's input and output.

Data Store:

A set of parallel lines shows a place for the collection of data items. A data store indicates that the data is stored which can be used at a later stage or by the other processes in a different order. They can be used to analyses an existing system or model of a new one. These are sources and destinations of the system's input and output.

Circle:

A circle (bubble) shows a process that transforms data inputs into data outputs. It is just like a function that changes the data, producing an output. It might perform computations for sort data based on logic or direct the dataflow based on business rules.

Data Flow:

A curved line shows the flow of data into or out of a process or data store. A dataflow represents a package of information flowing between two objects in the data-flow diagram, Data flows are used to model the flow of information into the system, out of the system and between the elements within the system.

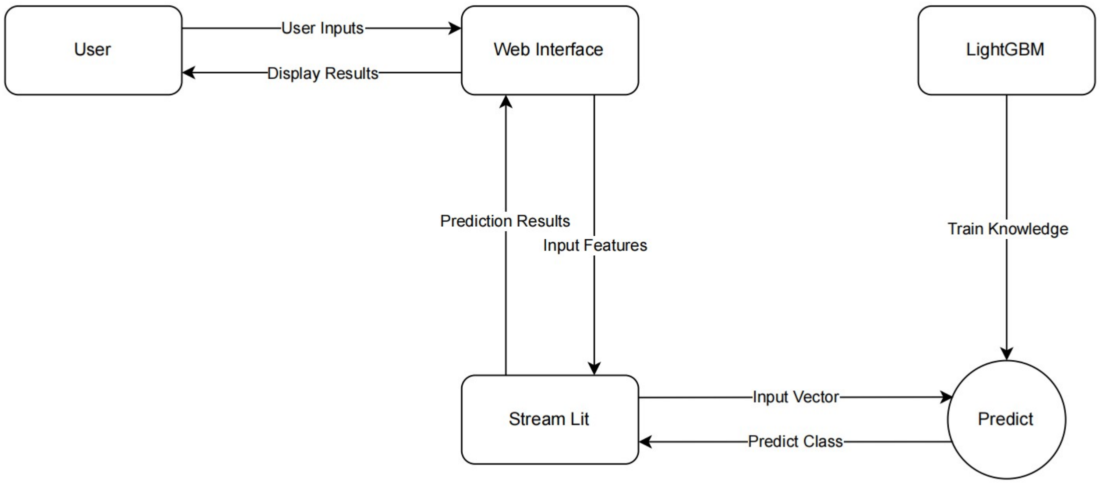


Figure 5.2 : Data Flow Diagram

5.3 Use Case Diagram:

Use Case during requirement elicitation and analysis to represent the functionality of the system. Use case describes a function by the system that yields a visible result for an actor. The actors are outside the boundary of the system, whereas the use cases are inside the boundary of the system. Use case describes the behavior of the system as seen from the actor's point of view.

The identification of actors and use cases result in the definitions of the boundary of the system i.e., differentiating the tasks accomplished by the system and the tasks accomplished by its environment. It describes the function provided by the system as a set of events that yield a visible result for the actor.

The purpose of use case diagram is to capture the dynamic aspect of a system. However, this definition is too generic to describe the purpose, as other four diagrams (activity, sequence, collaboration, and State chart) also have the same purpose. We will look into some specific purpose, which will distinguish it from other four diagrams. Use case diagrams are used to gather the requirements of a system including internal and external influences. These requirements are mostly design requirements.

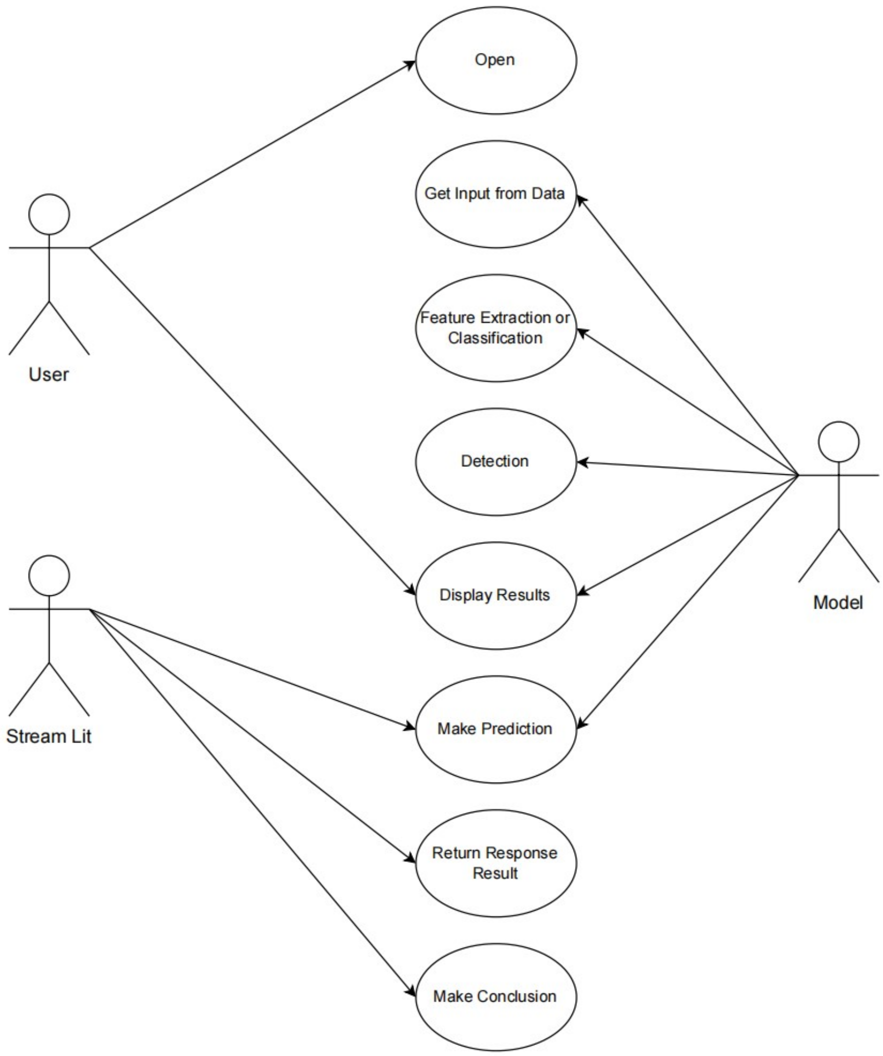


Figure 5.3 : Use Case Diagram of Web UI

5.4 System Architecture:

In software design, system architecture is the high-level structure of a software system. It defines the major components of the system, their relationships, and how they interact with each other. System architecture is important because it helps to ensure that the system is designed to meet its requirements and that it is efficient and scalable. System architecture in software design is typically created by software architects, who have a deep understanding of the system's requirements and the technologies that can be used to implement the system. Software architects work closely with other stakeholders, such as software developers, testers, and project managers, to ensure that the system architecture meets the needs of all stakeholders.

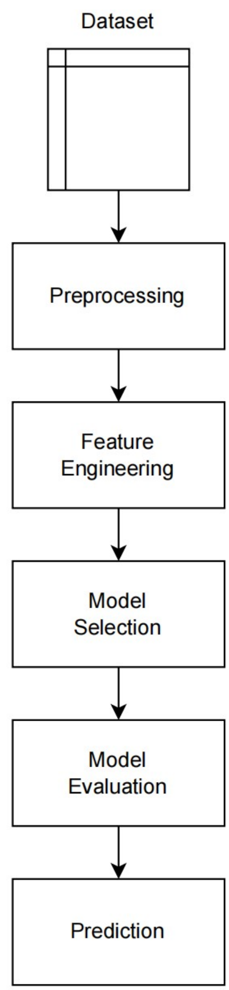


Figure 5.4 Pipeline of Disease Prediction

Chapter – 6

RESULTS & DISCUSSION

6. RESULTS & DISCUSSION

6.1 Evaluation Analysis:

In this project we have proposed a system for predicting Parkinson's disease using voice readings and the system predicts weather the person have Parkinson's disease or the person is healthy.

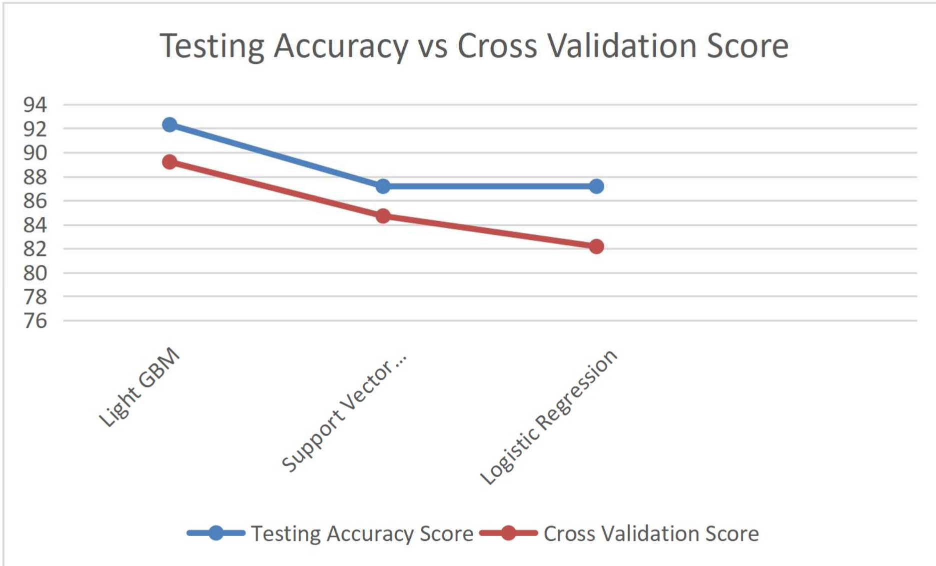
However while coming to the models build on Parkinson's disease detection, we used algorithms like LightGBM, Logistic Regression, Support vector Classifier. These results are shown in the following tables and graphs.

The results of these studies suggest that LightGBM is a promising algorithm for predicting Parkinson's disease. One of the advantages of using LightGBM for Parkinson's disease prediction is that it is a relatively fast and efficient algorithm. This makes it suitable for use in clinical settings, where rapid and accurate diagnosis is essential. Another advantage of using LightGBM is that it is able to learn from a variety of different data types, including demographic data, motor symptoms, non-motor symptoms, and voice recordings. This makes it possible to develop more comprehensive and accurate prediction models. Overall, LightGBM is a promising algorithm for predicting Parkinson's disease. It is important to further evaluate its performance on larger and more diverse datasets, but the results of the studies to date are encouraging.

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 6.1 : Performance of Models

Evaluation	**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



Figue 6.1 : 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%.

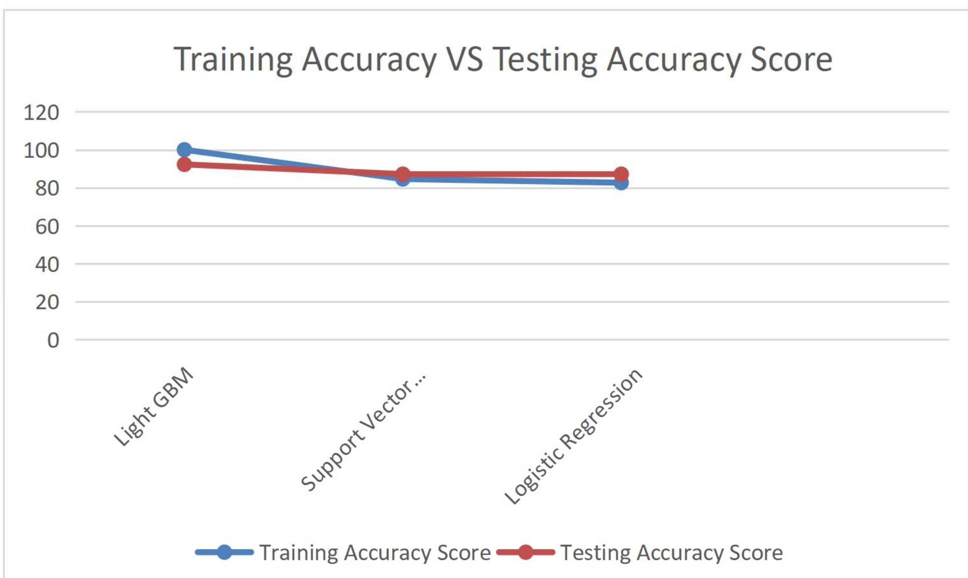


Figure 6.2 : Training Accuracy vs Testing Accuracy Score

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%.

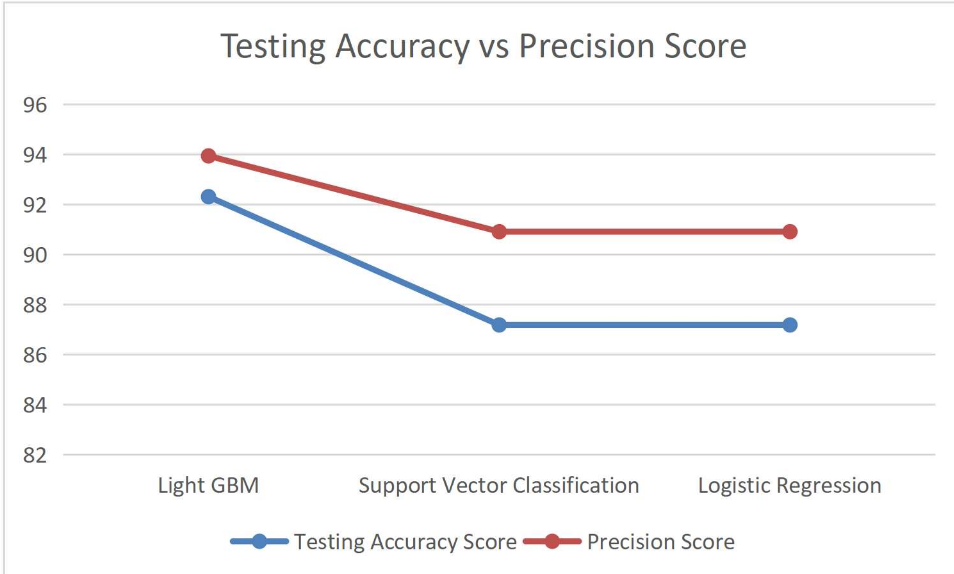


Figure 6.3 : Testing Accuracy vs Precision Score

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%.

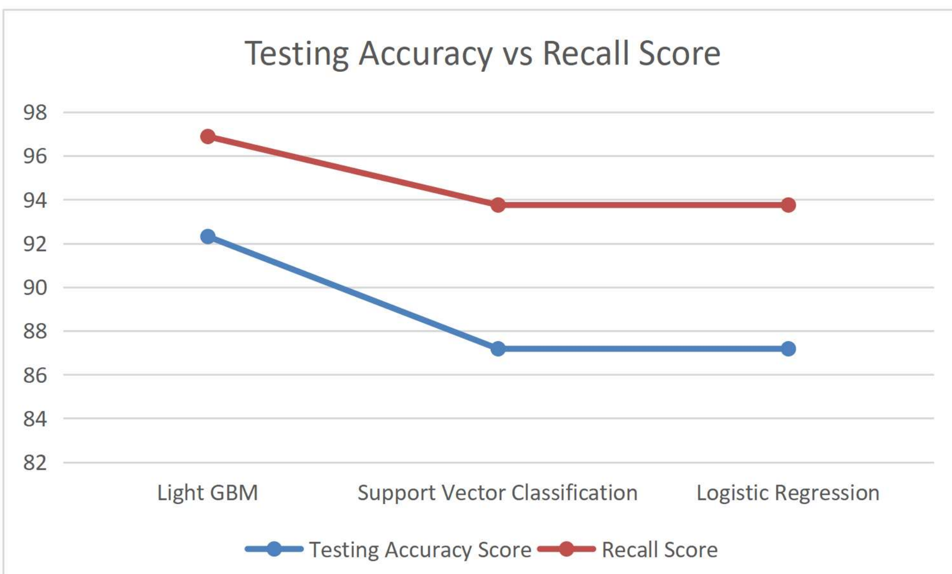


Figure 6.4 : Testing Accuracy vs Recall Score

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%.

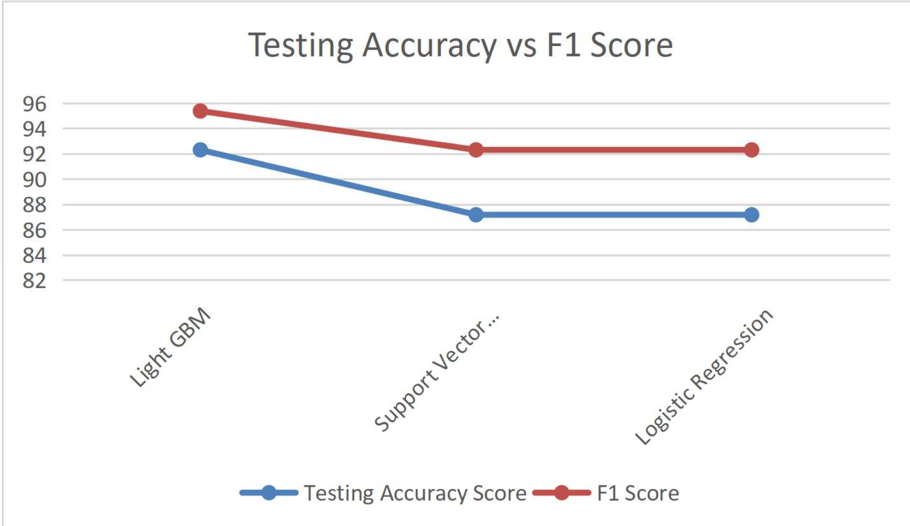


Figure 6.5 : Testing Accuracy vs F1 Score

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%.

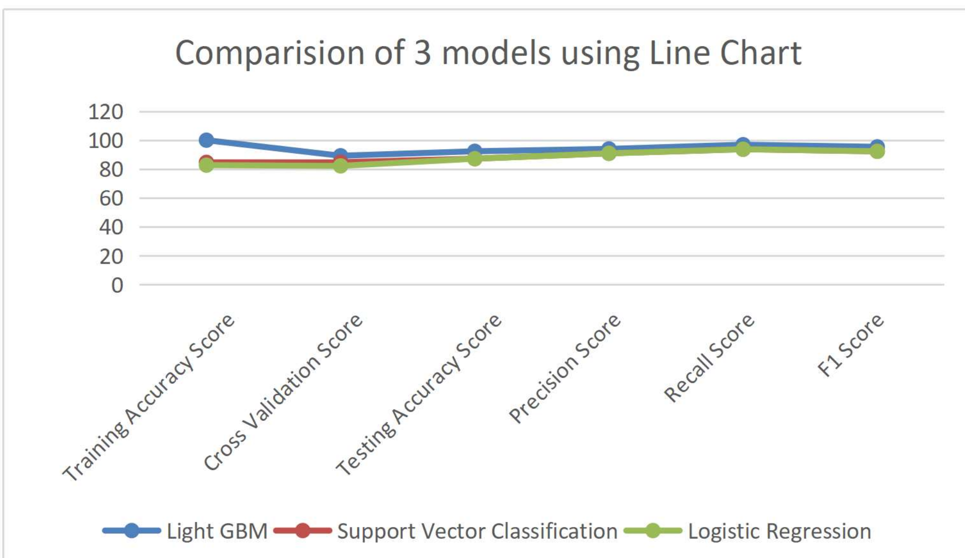


Figure 6.6 : Comparison of three models using line chart

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%.

6.2 Web Application User Interface Results :

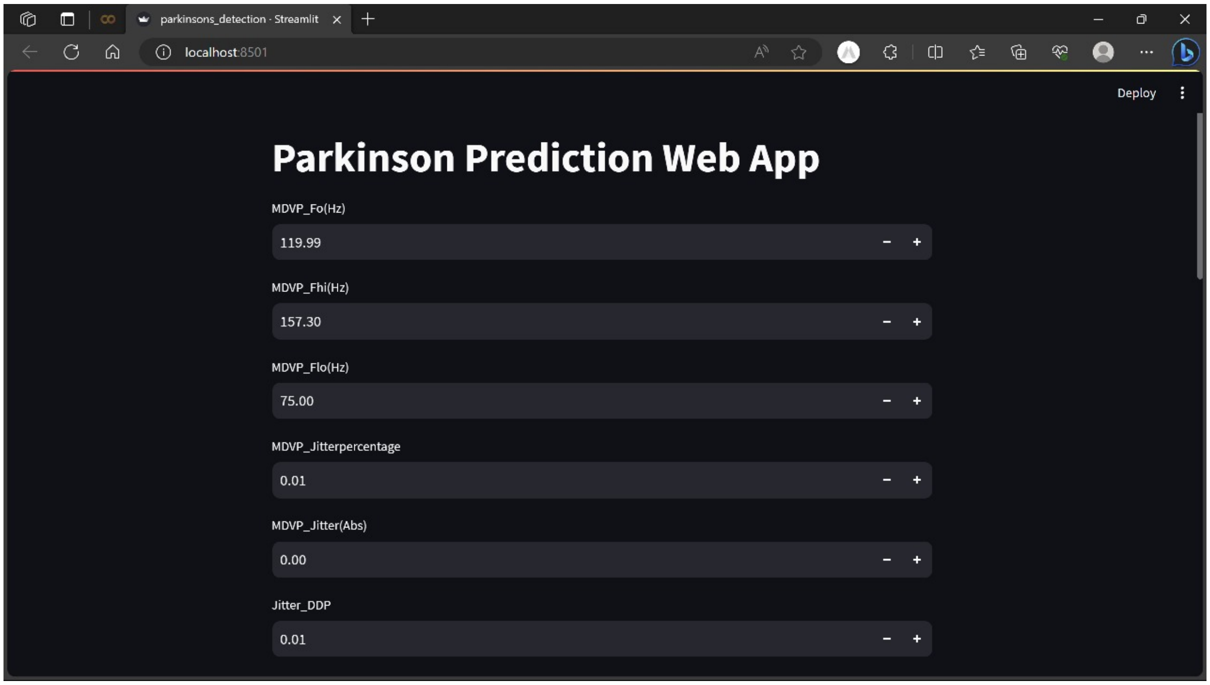


Figure 6.7 : Parkinson's disease prediction web page

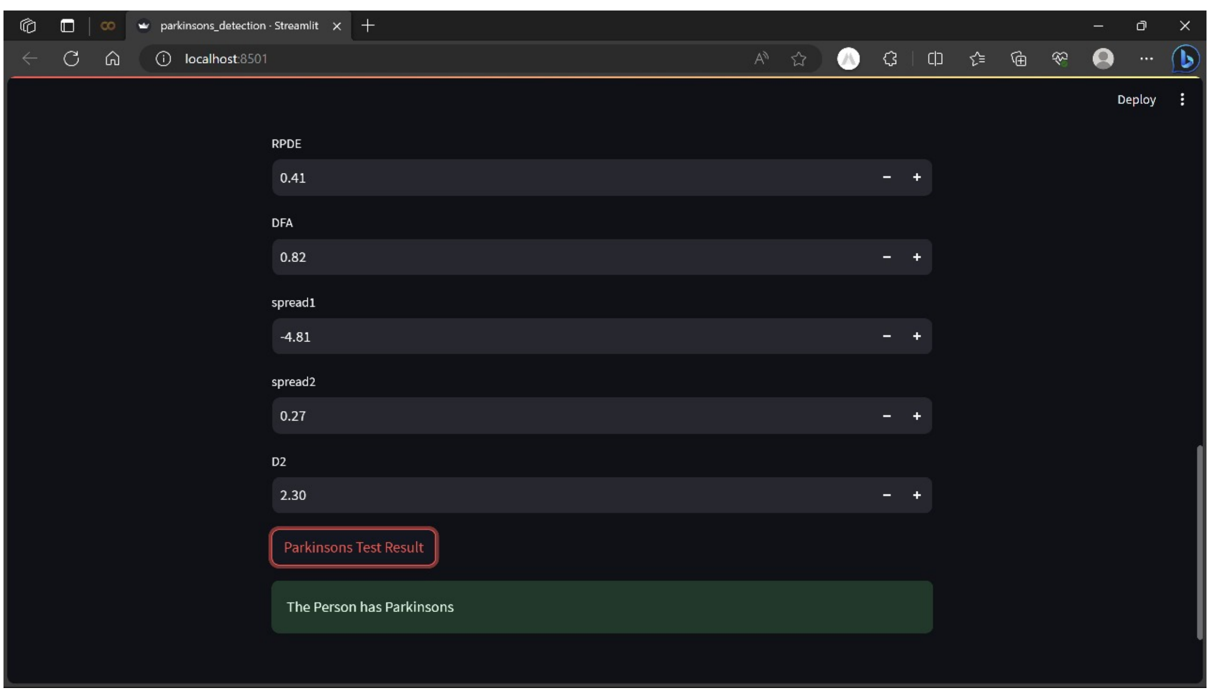


Figure 6.8 : Parkinson's disease prediction results page

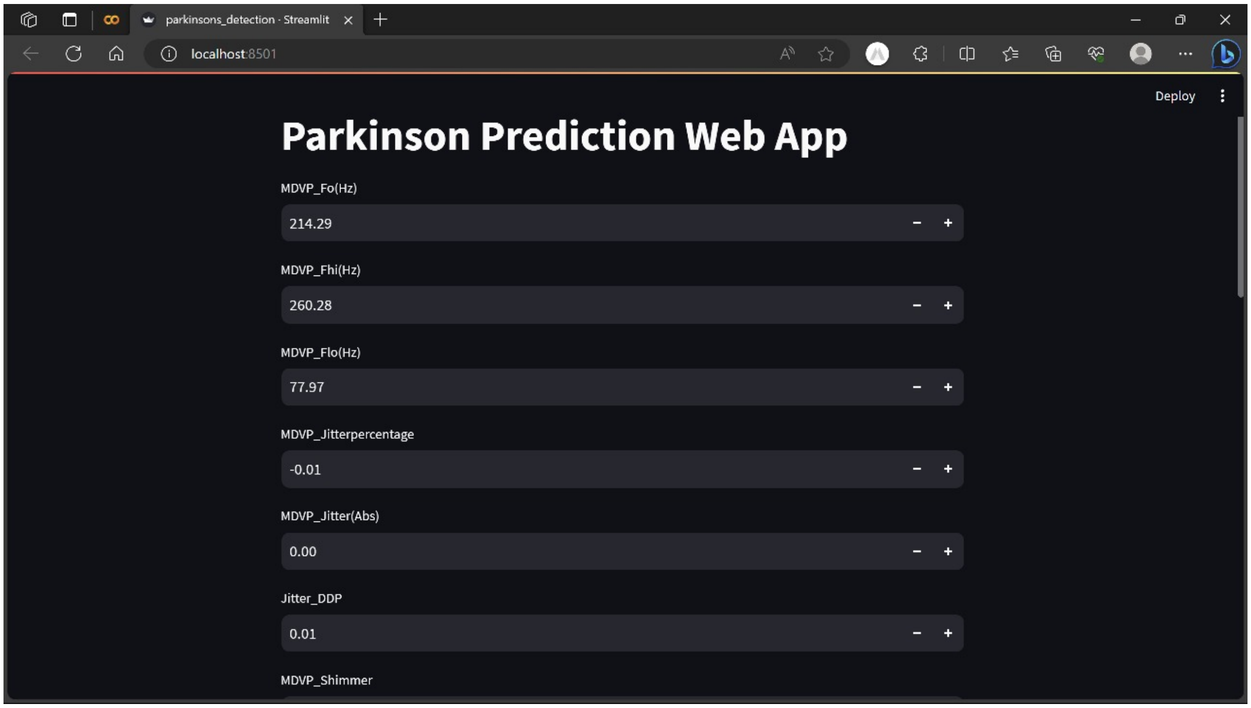
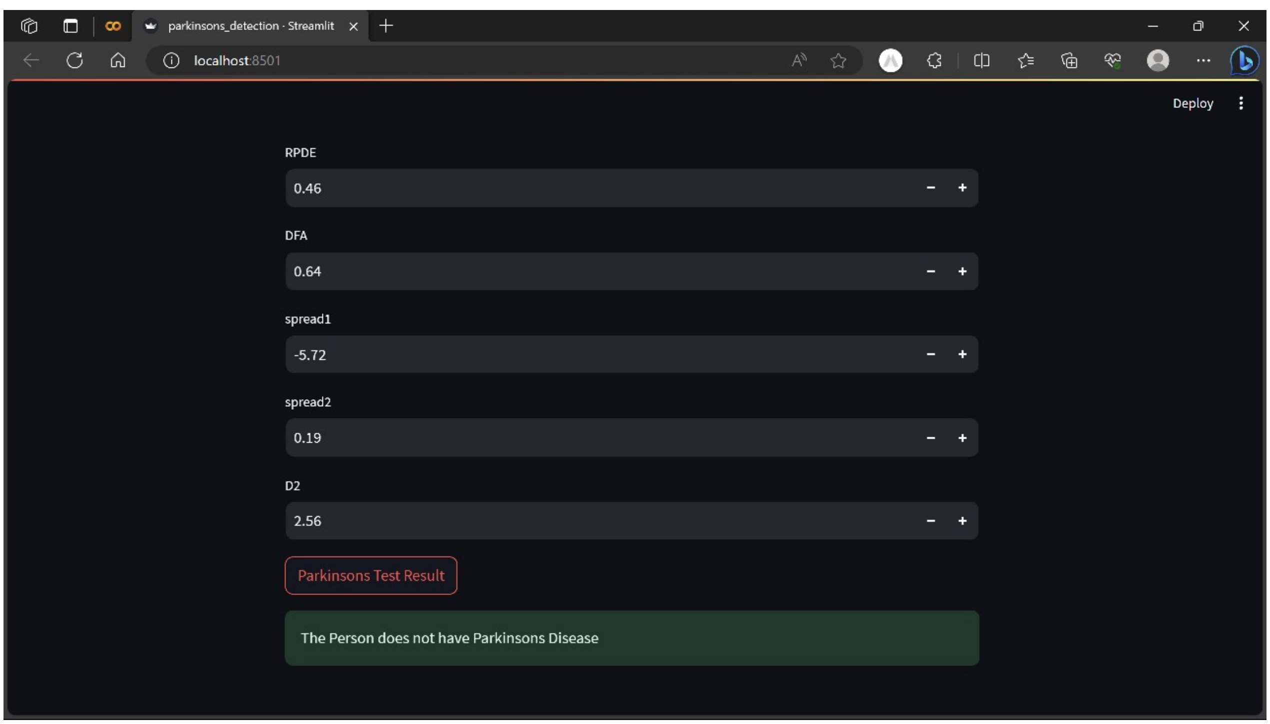


Figure 6.9 : Healthy person disease prediction web page



Fiigure 6.10 : Healthy person disease prediction results page

Chapter – 7

CONCLUSION & FUTURESCOPE

7. CONCLUSION & FUTURESCOPE

In this study, 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.

Chapter – 8

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.

[11] Nilashi, M., Abumalloh, R. A., Alyami, S., Alghamdi, A., & Alrizq, M. (2023). Parkinson's Disease Diagnosis Using Laplacian Score, Gaussian Process Regression and Self-Organizing Maps. Brain Sciences, 13(4), 543.

[12] Dimoudis, D., Tsolakis, N., Magga-Nteve, C., Meditskos, G., Vrochidis, S., & Kompatsiaris, I. (2023). InSEption: A Robust Mechanism for Predicting FoG Episodes in PD Patients. Electronics, 12(9), 2088.

APPENDIX - A

APPENDIX – A

```
pip install lightgbm
import lightgbm as lgb
from sklearn import svm
from sklearn.linear_model import LogisticRegression
import numpy as np
import pandas as pd
import random
from sklearn.model_selection import train_test_split
from sklearn.metrics import
accuracy_score,confusion_matrix,classification_report,f1_score,precision_score,recall_score
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
fp='C:\\Users\\Jagadeesh\\OneDrive\\Desktop\\Project\\Parkinson-disease-detection-
main\\Parkinson-disease-detection-main\\parkinsons.csv'
parkinson_data = pd.read_csv(fp)
parkinson_data.head().head()
parkinson_data.info()
parkinson_data.isna().sum()
parkinson_data.describe()
parkinson_data.columns
parkinson_data.duplicated().sum()
parkinson_data["status"].value_counts()
parkinson_data.hist(bins=50, figsize =(20,15), color = 'darkslategrey')
plt.show(block=False)
cor_matrix = parkinson_data.corr().abs()
plt.figure(figsize=(20,20))
print(cor_matrix)
sns.heatmap(cor_matrix,annot=True)
X = parkinson_data.drop(columns=['status','name'], axis=1)
Y = parkinson_data['status']
parkinson_data = pd.DataFrame(X)
```

```
print(parkinson_data.head())
df1 =parkinson_data.drop(columns=['MDVP_Fhi_Hz', 'MDVP_Fo_Hz', 'MDVP_Flo_Hz', 'HNR'],
axis=1)
print(); print(df1.head())
import numpy as np
from sklearn.model_selection import cross_val_score
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score,
confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
from termcolor import colored  # Import the colored function

def Evaluate_Performance(model, X_train, X_test, Y_train, Y_test):
    model.fit(X_train, Y_train)
    overall_score = cross_val_score(model, X_train, Y_train, cv=10)
    model_score = np.average(overall_score)
    Ypredicted = model.predict(X_test)
    print('-' * 80)
    print("\n • Training Accuracy Score : ", round(model.score(X_train, Y_train) * 100, 2))
    print(f" • Cross Validation Score : {round(model_score * 100, 2)}")
    print(colored(" ❖  Testing Accuracy Score : ", attrs=['bold']),
          colored(f"{round(accuracy_score(Y_test, Ypredicted) * 100, 2)}", color='black',
attrs=['bold']))
    print(' • Precision Score is :', round(precision_score(Y_test, Ypredicted) * 100, 2))
    print(' • Recall Score is :', round(recall_score(Y_test, Ypredicted) * 100, 2))
    print(' • F1-Score Score is :', round(f1_score(Y_test, Ypredicted) * 100, 2))
    print('-' * 80)
    conf_matrix = confusion_matrix(Y_test, Ypredicted)
    sns.heatmap(conf_matrix, annot=True, fmt='d', cmap=plt.cm.Blues, annot_kws={"size": 16})
    plt.title('Predicted Labels', y=1.05, fontsize=20, fontfamily='Times New Roman')
    plt.ylabel('True Labels', labelpad=25, fontsize=20, fontfamily='Times New Roman')
    plt.show()
pip install termcolor
X_train, X_test, Y_train, Y_test = train_test_split(df1, Y, test_size=0.2, random_state=7)
```

```
model = lgb.LGBMClassifier()
model.fit(X_train, Y_train)
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(Y_train, X_train_prediction)
Evaluate_Performance(model, X_train, X_test, Y_train, Y_test)
import pickle
filename = 'trained_modell.pkl'
pickle.dump(model, open(filename, 'wb'))
# loading the saved model
loaded_model = pickle.load(open('trained_modell.pkl', 'rb'))
```

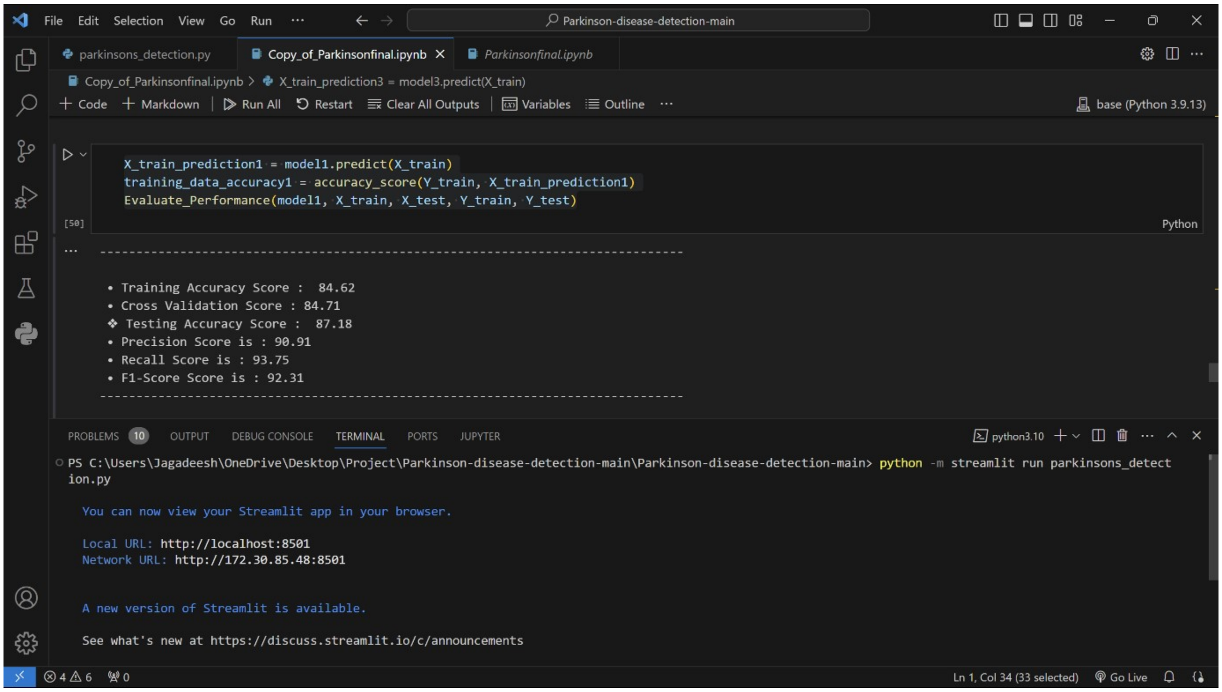
DATABASE CONNECTIVITY:

```
import numpy as np
import pickle
import streamlit as st
loaded_model=pickle.load(open('trained_modell.pkl', 'rb'))
def parkinson_prediction(input_data):
    # changing the input_data to numpy array
    input_data_as_numpy_array = np.asarray(input_data)
    # reshape the array as we are predicting for one instance
    input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)
    prediction = loaded_model.predict(input_data_reshaped)
    print(prediction)
    if (prediction[0] == 0):
      return 'The Person does not have Parkinsons Disease'
    else:
      return 'The Person has Parkinsons'
def main():
    # giving a title
    st.title('Parkinson Prediction Web App')
    # getting the input data from the user
    MDVP_Fo_Hz = st.number_input('MDVP_Fo(Hz)')
    MDVP_Fhi_Hz= st.number_input('MDVP_Fhi(Hz)')
```

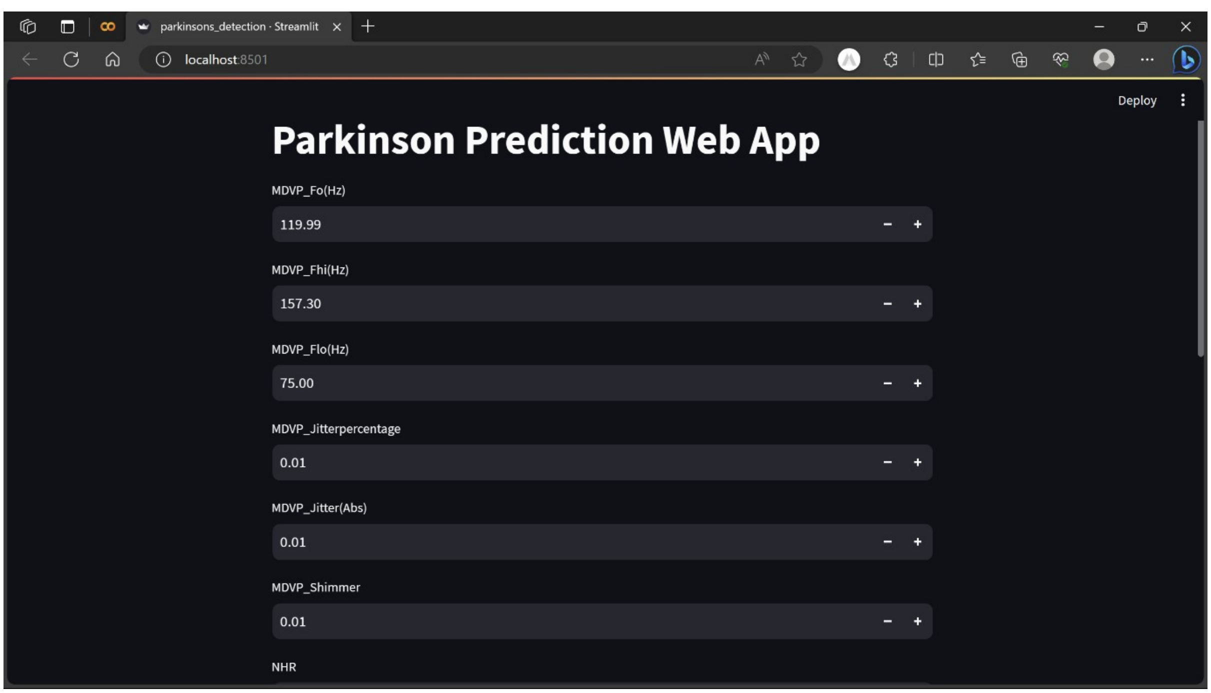
```
    MDVP_Flo_Hz= st.number_input('MDVP_Flo(Hz)')
    MDVP_Jitter_percentage= st.number_input('MDVP_Jitterpercentage')
    MDVP_Jitter_Abs= st.number_input('MDVP_Jitter(Abs)')
    Jitter_DDP=st.number_input('Jitter_DDP')
    MDVP_Shimmer= st.number_input('MDVP_Shimmer ')
    Shimmer_APQ3=st.number_input('Shimmer_APQ3')
    Shimmer_APQ5=st.number_input('Shimmer_APQ5')
    MDVP_APQ=st.number_input('MDVP_APQ')
    Shimmer_DDA=st.number_input(' Shimmer_DDA')
    NHR=st.number_input(' NHR ')
    HNR= st.number_input('HNR')
    RPDE=st.number_input(' RPDE ')
    DFA=st.number_input(' DFA ')
    spread1 = st.number_input(' spread1 ')
    spread2 = st.number_input(' spread2 ')
    D2=st.number_input(' D2')

    # code for Prediction
    diagnosis = ''
    # creating a button for Prediction
    if st.button('Parkinsons Test Result'):
        diagnosis=
parkinson_prediction([[MDVP_Fo_Hz,MDVP_Fhi_Hz,MDVP_Flo_Hz,MDVP_Jitter_percentage,
MDVP_Jitter_Abs,Jitter_DDP,MDVP_Shimmer,Shimmer_APQ3,Shimmer_APQ5,MDVP_APQ,
Shimmer_DDA,NHR,HNR ,RPDE,DFA,spread1,spread2,D2]])
    st.success(diagnosis)
```

APPENDIX - B

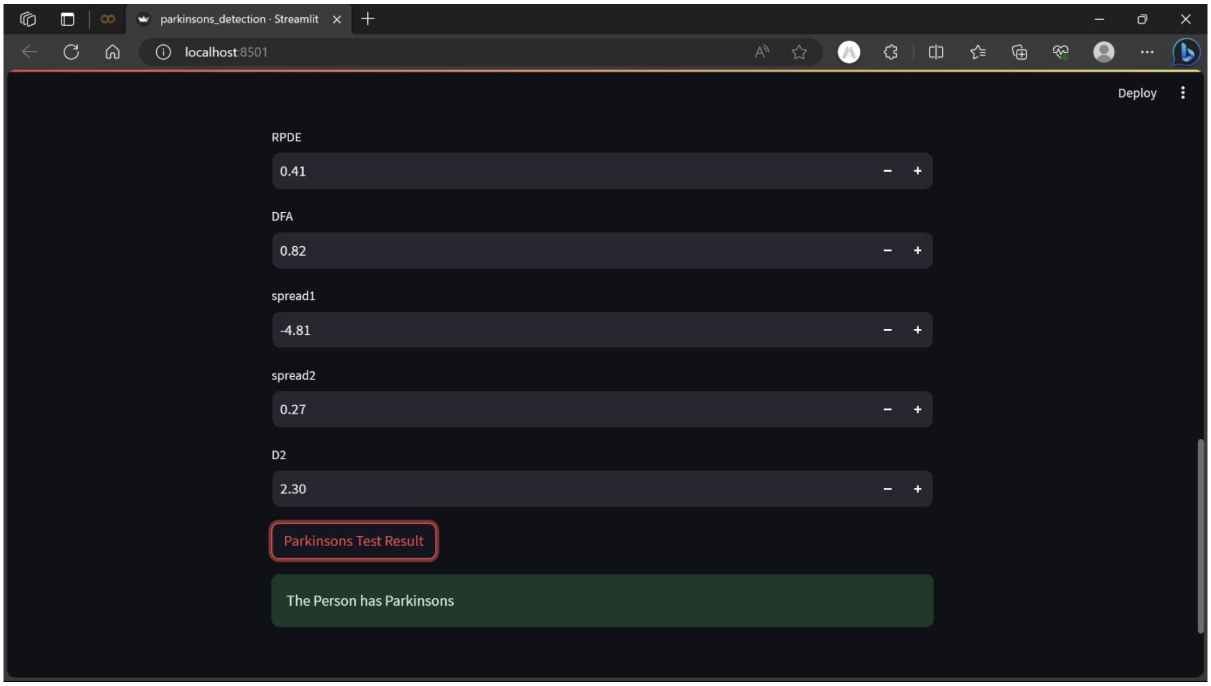
APPENDIX – B

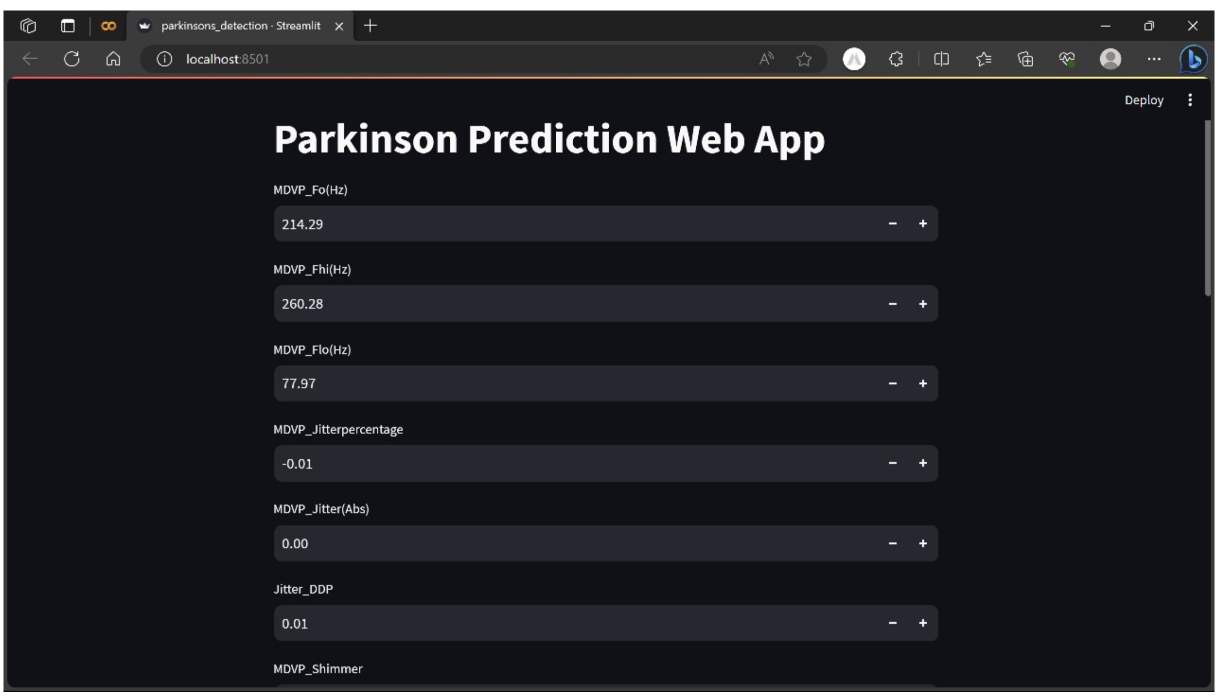


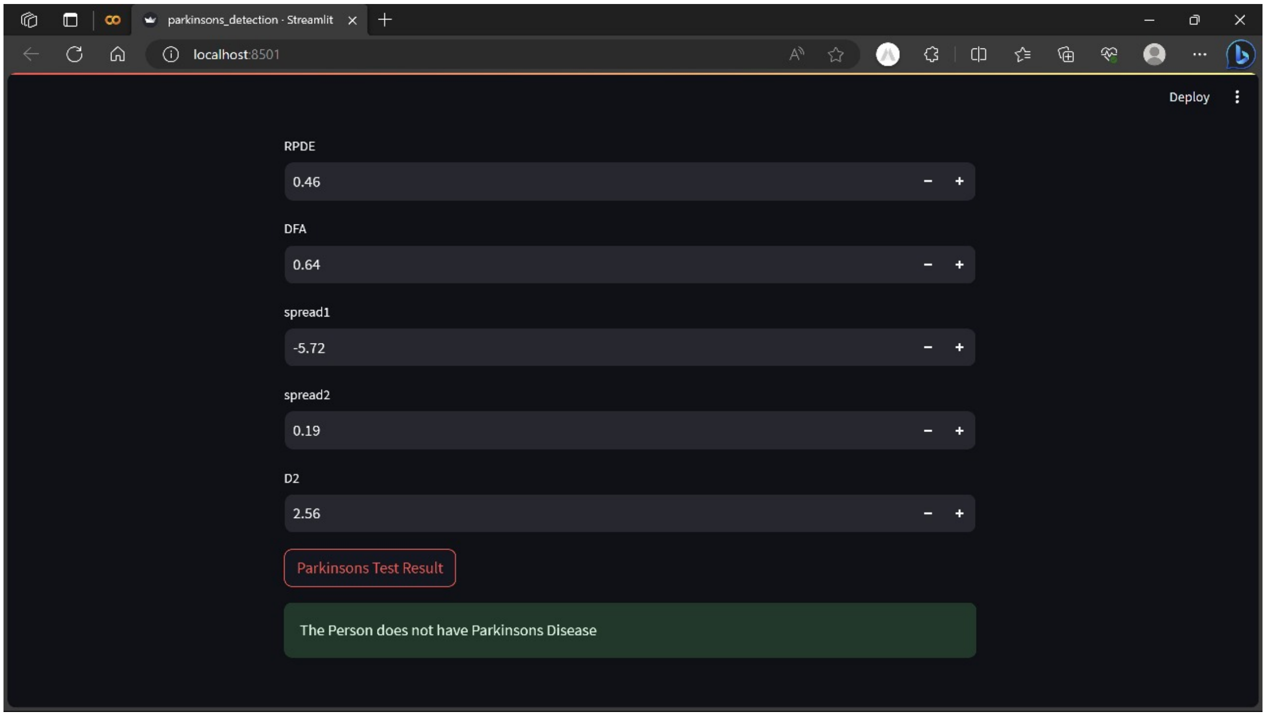
Parkinson-disease-detection-main
parkinsons_detection.py
Copy_of_Parkinsonfinal.ipynb
Parkinsonfinal.ipynb
X_train_prediction1 = model1.predict(X_train)
training_data_accuracy1 = accuracy_score(Y_train, X_train_prediction1)
Evaluate_Performance(model1, X_train, X_test, Y_train, Y_test)
• Training Accuracy Score : 84.62
• Cross Validation Score : 84.71
• Testing Accuracy Score : 87.18
• Precision Score is : 90.91
• Recall Score is : 93.75
• F1-Score Score is : 92.31
PS C:\Users\Jagadeesh\OneDrive\Desktop\Project\Parkinson-disease-detection-main\Parkinson-disease-detection-main> python -m streamlit run parkinsons_detection.py
You can now view your Streamlit app in your browser.
Local URL: http://localhost:8501
Network URL: http://172.30.85.48:8501
A new version of Streamlit is available.
See what's new at https://discuss.streamlit.io/c/announcements



parkinsons_detection · Streamlit
localhost:8501
Deploy
Parkinson Prediction Web App
MDVP_Fo(Hz)
119.99
MDVP_Fhi(Hz)
157.30
MDVP_Flo(Hz)
75.00
MDVP_Jitterpercentage
0.01
MDVP_Jitter(Abs)
0.01
MDVP_Shimmer
0.01
NHR


parkinsons_detection · Streamlit
localhost:8501
Deploy
RPDE
0.41
DFA
0.82
spread1
-4.81
spread2
0.27
D2
2.30
Parkinsons Test Result
The Person has Parkinsons


parkinsons_detection · Streamlit
localhost:8501
Deploy
Parkinson Prediction Web App
MDVP_Fo(Hz)
214.29
MDVP_Fhi(Hz)
260.28
MDVP_Flo(Hz)
77.97
MDVP_Jitterpercentage
-0.01
MDVP_Jitter(Abs)
0.00
Jitter_DDP
0.01
MDVP_Shimmer


parkinsons_detection · Streamlit
localhost:8501
Deploy
RPDE
0.46
DFA
0.64
spread1
-5.72
spread2
0.19
D2
2.56
Parkinsons Test Result
The Person does not have Parkinsons Disease

APPENDIX - C

IEEE Technically Sponsored International Conference on Energy, Materials and Communication Engineering (ICEMCE) 2023 : Submission (110) has been created.

From: Microsoft CMT <email@msr-cmt.org>
Sent: Thursday, September 7, 2023 3:22:06 PM
To: Anilkumar Ch <anilkumar.ch@gmrit.edu.in>
Subject: IEEE Technically Sponsored International Conference on Energy, Materials and Communication Engineering (ICEMCE) 2023 : Submission (110) has been created.

Hello,

The following submission has been created.

Track Name: ICEMCE2023

Paper ID: 110

Paper Title: A Spatial Graphical Convolution Neural Network to Denoising Auto-encoder Freezing of Gait in Parkinson’s Disease

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). Early detection and treatment of freezing of gait (FOG) can significantly enhance a person with Parkinson's disease's quality of life. 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. High levels of specificity, sensitivity, and accuracy were attained by the model.

Created on: Thu, 07 Sep 2023 09:52:04 GMT

Last Modified: Thu, 07 Sep 2023 09:52:04 GMT

Predictive Modeling of Parkinson’s Disease Progression using LightGBM Classifier

Ch. Anil Kumar
Information Technology
GMR Institute Of Technology
Rajam, Andhra Pradesh, India
anilkumar.ch@gmrit.edu.in

D. Shamitha
Information Technology
GMR Institute Of Technology
Rajam, Andhra Pradesh, India
20341A1221@gmrit.edu.in

K. Purna Chandu
Information Technology
GMR Institute Of Technology
Rajam, Andhra Pradesh, India
20341A1245@gmrit.edu.in

G. Avinash
Information Technology
GMR Institute Of Technology
Rajam, Andhra Pradesh, India
20341A1224@gmrit.edu.in

D. Pavan Kumar
Information Technology
GMR Institute Of Technology
Rajam, Andhra Pradesh, India
20341A1222@gmrit.edu.in

***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). Early detection and treatment of freezing of gait (FOG) can significantly enhance a person with Parkinson's disease's quality of life. 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. High levels of specificity, sensitivity, and accuracy were attained by the model.**

Index Terms - Freezing of gait (FOG), LightGBM, Motor and non-motor symptoms, Gait patterns, Feature engineering, Gradient boosting.

I. 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. The degeneration of these dopamine-producing neurons is one of the defining signs of Parkinson's disease. 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 trapped in 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 number has 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. Those affected 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.

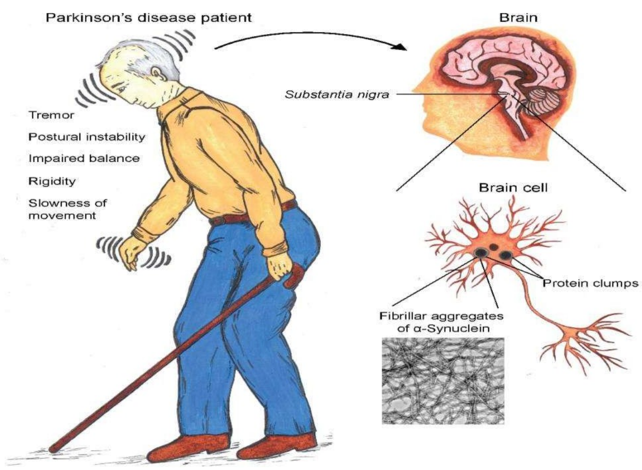


Figure: Parkinson's Disease Symptoms.

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 ofParkinson's disease with early onset have been connected to synuclein protein mutations. The development and progression 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.

II. Related Work

[1] Multipliers Esfahani, et.al. 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 avoidfalls. Wearable technology is accessible, small, and has a long battery life; all users need to doto 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.

[2] Noor, et.al. 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. Toget 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.

[3] Lin, C. H et al. 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.

[4] Wang, et.al. 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%.

[5] Kwon et al.The proposed method in the paper is a spatial-temporal graphical convolutional 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 Parkinsonian 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 Parkinsonian patients.

[6] Borzì et al. 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.

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

[8] Zhang, et al. This study proposes a novel method for recognizing Parkinsonian gait from forward videos. The suggested technique, known as WM-STGCN, uses a weighted adjacency matrix with virtual connections and multi-scale temporal convolution. It is a spatiotemporal graph convolutional network. 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. With an F1 score of 92.85%, WM-STGCN surpassed baseline approaches like LSTM, KNN, decision tree, AdaBoost, and ST-GCN. Its accuracy rate was 87.1%.

[9] Cai, et al. 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.

[10] Kour et al. 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.

III. PROPOSED METHODOLOGY

Microsoft first unveiled the LightGBM gradient boosting framework a few years ago, and it has since grown to be a significant force in the gradient boosting tool market. LightGBM is six times faster than XGBoost, and it offers amazing speed. This framework solves the drawbacks of the histogram-based approach that underlies all Gradient Boosting Decision Tree (GBDT) systems. It does so by utilizing tree-based learning techniques. The LightGBM Algorithm combines two methods known as GOSS and EFB, which together give it an edge over other GBDT frameworks and permit it to perform well.

It uses two novel techniques:

Gradient-based One Side Sampling(GOSS)

Exclusive Feature Bundling (EFB)

Gradient-based One Side Sampling Technique for LightGBM:

The information gain calculation involves several roles for various data instances. The knowledge benefit will be greater for the situations with larger gradients To ensure the accuracy of information gain estimation, GOSS functions by systematically removing samples with small gradients. It keeps cases with large gradients, usually those that are greater than a certain threshold or rank highly in the gradient distribution. When the information gain varies greatly in magnitude, this method can result in a gain estimation that is more precise than one that uses uniformly random sampling.

Exclusive Feature Bundling Technique for LightGBM:

There is a chance to create an effective strategy for drastically lowering the number of features without suffering considerable information loss given the common issue of having limited high-dimensional data. More specifically, because they never exhibit non-zero values concurrently in sparse feature spaces, many features are incompatible with one another. One feature, known as a "Exclusive Feature Bundle," can be created by securely combining the unique features.

Architecture of LightBGM:

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 ischosen to grow.

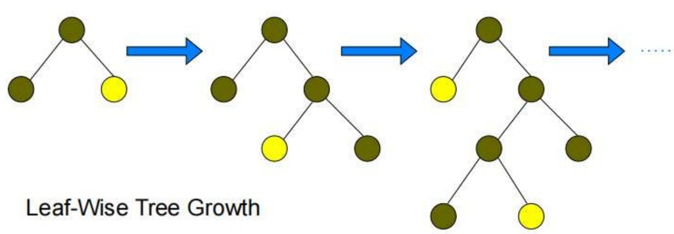


Figure: Light GBM Architecture

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 terminationcriteria. 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.

The following benefits are built into its efficient distribution design:

- Greater speed of training as well as greater efficiency.
- Less memory is used.
- Better precision.
- Support for GPU-accelerated parallel learning.
- capable of managing massive amounts of data.

IV. CONCLUSION

In this study, 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.

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.

20IT701 Project Work

0 0 16 8

At the end of the project work the students will be able to

1. Identify a contemporary engineering application to serve the society at large
2. Use engineering concepts and computational tools to get the desired solution
3. Justify the assembled/fabricated/developed products intended.
4. Organize documents and present the project report articulating the applications of the concepts and ideas coherently
5. Demonstrate ethical and professional attributes during the project implementation.
6. Execute the project in a collaborative environment.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	3	2				3	2						3	3
CO2	3	3			3								3	3
CO3	3	3	3	2							2		3	3
CO4										3		2	3	3
CO5								3					3	3
CO6									3				3	3