PARKINSON’S DISEASE DETECTION USING SUPPORT VECTOR MACHINE ALGORITHM

A PROJECT REPORT

Submitted by

JAASIR AHAMED A	**(311820205011)**
MOHAMED ASIF H	**(311820205018)**
SYED ALI S	**(311820205047)**

In partial fulfillment for the award of the degree
of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

MOHAMED SATHAK A.J. COLLEGE OF ENGINEERING
SIRUSERI IT PARK, OMR, CHENNAI - 603 103

ANNA UNIVERSITY : CHENNAI 600 025

MAY 2024

ANNA UNIVERSITY : CHENNAI 600 025

BONAFIDECERTIFICATE

It is certified that this project report titled**“Parkinson’s Disease Detection Using Support Vector Machine Algorithm”** is the bonafide work of **"JAASIR AHAMED A(311820205011), MOHAMED ASIF H(311820205018), SYED ALI S(311820205047)"**who carried out the project work under my supervision**.**

SIGNATURE
Dr.D.PRAKASH,M.E.,M.Tech,.Ph.D.
HEAD OF THE DEPARTMENT,
ASSOCIATE PROFESSOR
Department of Information Technology
Mohamed Sathak AJ College of Engineering
IT Sipcot, OMR,Siruseri
Chennai-603 103

SIGNATURE
Mrs.M.JEBAMALAR,M.E.
SUPERVISOR,
ASSISTANT PROFESSOR
Department of InformationTechnology
Mohamed Sathak AJ College of Engineering
IT Sipcot, OMR,Siruseri
Chennai-603 103

Submitted for the university practical examination held at **MOHAMED SATHAK A.J COLLEGE OF ENGINEERING, CHENNAI on** ……………………

INTERNALEXAMINER **EXTERNALEXAMINER**

ABSTRACT

Parkinson's disease, a progressive neurological disorder, often goes undetected in its initialstagesduetothesubtleandeasilyoverlookednatureofitsearlysymptoms.This neural disease affects the central nervous system, manifesting through tremors, stiffness, changes in facial expressions, and fever. Even medical professionals sometimes struggle to identify Parkinson's in its early phases, leading to delayed diagnosis and treatment, which can exacerbate the condition's effects.

Our project aims to address this critical challenge by developing a highly accurate Parkinson's disease detection system capable of identifying the condition from its earliest stages. We have trained our system on a comprehensive dataset, enabling it to recognize the subtle patterns and biomarkers associated with Parkinson's. By leveraging advanced machine learning techniques and analyzing various physiological signals, our system can provide early detection with high precision.

Early intervention is crucial for managing Parkinson's disease, as it can significantly slow its progression and improve the quality of life for those affected. Our data-driven approach, combined with cutting-edge algorithms, represents a significant step forward in the fight against this debilitating condition. Early detection empowers individuals to seek timely medical attention and appropriate treatment, offering hope and better outcomes for those at risk of Parkinson's disease.

Through our project, we aim to raise awareness about the importance of early detection and provide are liable tool to identify Parkinson's disease before it progresses to more advanced stages, ultimately improving the lives of those affected by this neurological disorder.

Keywords: Parkinson's disease, early detection, machine learning, physiological signals, data-driven, algorithms, early intervention, quality of life, neurological disorder.

ACKNOWLEDGEMENT

Firstly, we express our gratitude to the almighty for facilitating the successful completion of this project under all circumstances.

We extend our sincere and heartfelt appreciation to **ALHAJ JANAB YOUSUF S.M., Chairman, Mr. MOHAMED SATHAK, Director, JANAB A S.H. SHARMILA, Secretary, and JANAB P.R.L HAMID IBRAHIM, Executive Director**, Mohamed Sathak AJ College of Engineering, Chennai.

We take immense pleasure in conveying our heartfelt thanks to our college **Principal, Dr. K.S. SRINIVASAN, Ph.D.,** Mohamed Sathak A J College of Engineering, for granting us permission to undertake this project.

We are also grateful to our Head of the Department, **Dr.D.PRAKASH, M.E., M.Tech., Ph.D., Associate Professor**, Mohamed Sathak A J College of Engineering, and our Supervisor, **Mrs M JEBAMALAR M.E., Assistant Professor**, Mohamed Sathak A J College of Engineering, for their constant guidance and motivation during this project.

We extend our heartfelt appreciation to **Dr. D. WESLIN, M.E., Ph.D.,** from **Mohamed Sathak A J College of Engineering**, acknowledging his pivotal role as the **RESEARCH HEAD** whose unwavering guidance and inspiration fueled our project's success.

We would also like to express our appreciation to all the **Faculty Members and Non-teaching Staff** of the Department of Information Technology for their valuable support.

Finally, we would like to acknowledge the unwavering support of our Parents and Friends throughout the duration of this project.

TABLE OF CONTENTS

CHAPTER NO. TITLE PAGE NO.

LIST OF FIGURES

CHAPTER 1

1 INTRODUCTION:

Parkinson's Disease (PD) is a complex neurodegenerative disorder affecting millions globally, with its prevalence increasing as the population ages. Early diagnosis remains a significant challenge due to the gradual onset of symptoms. Early detection iscrucialfor timely treatment and interventions, potentially slowing disease progression andimproving quality of life.

Traditional diagnostic methods heavily rely on clinical evaluations, which may not accurately detect PD in its early stages. This has led researchers to explore alternative approaches, including analyzing non-motor symptoms like voice and speech impairments. Evidence suggests individuals with PD exhibit distinct vocal patterns, including changes in fundamental frequency, jitter, shimmer,and other acoustic measures, even before motor symptom onset.

Our methodology involved rigorous data collection and preprocessing, obtaining clinicalvoicesamplesandmeticulouslyextractingrelevantacousticfeaturesusing state-of-the-art signal processing techniques.These features were carefully selected based on their associations with PD and potential to discriminate between healthy and affected individuals. Through extensive experimentation and validation, we aimed to develop an accurate and robust model capable of distinguishing individuals with and without PD based solely on their voice characteristics.

The successful development of such a diagnostic model holds the potential to revolutionize early PD detection and management. By enabling earlier interventions and personalized treatment plans, our research could significantly improve patient outcomes and alleviate the substantial burden imposed by this debilitating condition worldwide.

1.1. PROBLEMSTATEMENT:

Parkinson's Disease (PD) is a progressiveneurodegenerativedisordercharacterizedby motor and non-motor symptoms, affecting millions of individuals worldwide. Early and accurate diagnosis of PD is crucial for initiating timely treatment and interventions, potentially slowing disease progression and improving the quality of life for those affected. However, traditional diagnosticmethodsheavilyrelyonclinicalevaluationsand neurological examinations, which may not always accurately detect the disease in itsearly stages when symptoms are subtle.

Recent studies have shown that individuals with PD exhibit distinct patterns in their vocal characteristics, including changes in fundamental frequency, jitter, shimmer, and other acoustic measures, even before the onset of motor symptoms. These voice-related featureshaveemergedaspotentialbiomarkersforearlyPDdetection.However,manually analyzing andinterpretingthesecomplexacousticmeasurescanbechallengingandprone to subjective biases.

The objective of this project is to develop a robust and accurate diagnostic model for Parkinson'sDiseasebyleveragingmachinelearning techniques and the analysis of voice-related features. Specifically, we aim to employ the Support Vector Machine (SVM) algorithm, a powerful supervised learning method, to classify individuals as having or not having PD based on their vocal characteristics.

By training the SVM model on a comprehensive dataset of acoustic measuresextracted from high-qualityvoicerecordingsofindividualswithandwithoutPD,weseek to identify patterns and relationships that can effectively distinguish between the two groups.Thesuccessfuldevelopmentofsuchadiagnosticmodelholdsthepotentialto revolutionizeearlyPDdetection,enablingearlierinterventions,personalizedtreatment plans, and improved patient outcomes.

Furthermore, the project aims to address the limitations of traditional diagnostic methods by providing an objective, data-driven approach toPDdiagnosis,leveragingthe power of machine learning algorithms and advanced signal processing techniques. The ultimate goal is to contribute to the field of Parkinson's Disease research and healthcare by developing a reliableandnon-invasivetoolforearlydiseasedetection,pavingtheway for more effective disease management andalleviatingthesubstantialburdenimposedby this debilitating condition on individuals, families, and healthcare systems worldwide.

1.2. SIGNIFICANCEANDPURPOSE:

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects millions of individuals globally, with a significant impact on their quality of life and independence. As the world's population continues to age, the prevalence of PD is expected to rise, highlighting the urgent need for effective diagnostic tools and early interventions. Early and accurate diagnosis is crucial for initiating timely treatment and management strategies, which can potentially slow the progression of the disease and alleviate its debilitating symptoms.

Despite advancements in medicalresearch,traditionaldiagnosticmethodsforPD,such as clinical evaluations and neurological examinations, often failtodetectthediseaseinits early stages when symptoms are subtle and easilyoverlooked.Thisdelayindiagnosiscan lead to missed opportunities for early interventions, potentially exacerbating the disease's impact on individuals and their families.

Theanalysisofvoice-relatedfeatureshasemergedasapromisingavenueforearly PD detection, as individuals with the disease exhibit distinct patterns in their vocal characteristics, includingchangesinfundamentalfrequency,jitter,andshimmer.However, manually analyzing these complex acoustic measures can be challenging and prone to subjective biases, hindering the widespread adoption of this approach in clinical settings.

The primary purpose of this project is to develop a robust and accurate diagnostic model for Parkinson's Disease byleveragingmachinelearningtechniquesandtheanalysis of voice-related features. By employing the powerful Support Vector Machine (SVM) algorithm, a supervised learning method renowned for its ability to classify data effectively, we aim to distinguish individuals with and without PD based on their vocal characteristics.

Through rigorous data collection and preprocessing, we will obtain high-quality voice recordings from individuals with and without PD, and meticulously extract relevant acoustic features using advanced signal processing techniques. These features will be carefully selected based on their established associations with PD and their potential to discriminate between healthy and affected individuals.

By training theSVMmodelonthiscomprehensivedataset,weseektoidentifypatterns and relationships that can reliably differentiate between individuals with early-stage PD and those without the disease. The successful development of such a diagnostic model holds the potential to revolutionize early PD detection, enabling timely interventions, personalized treatment plans, and ultimately, improved patient outcomes.

1.3. SCOPE:

The scope of this project encompasses the development and evaluation of an improved machine learning-based diagnostic model for Parkinson's Disease (PD) using the Support Vector Machine (SVM) algorithm. The project aims to leverage the power of a novel algorithm designed to enhance the accuracy and performance of the SVM technique in classifying individuals as having or not having PD based on their relevant features.

The project will involve the following key components:

1.Data Collection and Preprocessing: Obtain a comprehensive dataset containing relevantfeaturesandmedicalinformationfromindividualswithandwithoutParkinson's Disease, ensuring a diverse and representative sample.Implement rigorous data preprocessing techniques, including handling missing data, normalization, and quality checks, to ensure the integrity and reliability of the dataset.

2.Feature Engineering and Selection: Perform exploratory data analysis to identify the most informative features for the PD classification task.Apply feature engineering techniques,ifnecessary,toderivenewfeaturesortransformexistingonestobettercapture the underlying patterns associated with PD.Conduct feature selection and dimensionality reduction to eliminate redundant or irrelevant features, improving model efficiency and performance.

3.Algorithm Development: Design and implement a novel algorithm that enhances the accuracyandperformanceoftheSVMalgorithmforthePDclassificationtask.Incorporate domain-specific knowledge or expert insights to guide the algorithm's development and optimization.Explore different kernel functions, hyperparameter tuning, and ensemble techniques to further improve the model's performance.

4.ModelTrainingandValidation: Splitthedatasetintotrainingandtestingsets,ensuring appropriate cross-validation techniques to evaluate the model's generalization capabilities.TraintheenhancedSVMmodelusingthedevelopedalgorithmandoptimized hyperparameters.Validate the model's performance on unseen test data to assess its real-worldapplicabilityandrobustness.

5.ModelEvaluationandComparison: Conductcomprehensiveevaluationofthetrained model using various performance metrics, such as accuracy, precision, recall, and F1-score.Perform comparative analyses with the standard SVM algorithm and other machinelearningmodelstobenchmarktheperformanceimprovementsachievedbythe developed algorithm.Investigate the potential impact of factors such as age, gender, or disease severity on the model's performance.

6.ResultAnalysisandInterpretation: Analyzeandinterprettheresultsobtainedfromthe enhanced SVM model, identifying the contributions of the novel algorithm and its impact on the model's accuracy and performance.Explore the model's limitations and potential areas for further improvement or future research.

7.Deployment and Integration: Investigate the potential for deploying the developed modelinclinicalsettingsorasascreeningtoolforPDdetection.Exploreopportunitiesfor integrating the model with existing healthcare systems or mobile applications for widespread accessibility.

CHAPTER 2

2.LITERATUREREVIEW:

In recent years, machine learning techniques have gained significant attention in the field of PD diagnosis due totheirabilitytoanalyzecomplexdatasetsandidentifypatterns that may not be readily apparent to human observers. Among the various machinelearning algorithms, the Support Vector Machine (SVM) has been widely exploredforits potential in classifying individuals with and without PD based on various features.

Several studies have investigated the use ofSVMforPDdiagnosis.Forinstance,Sakar et al. (2019) employed an SVM model to classify individuals with PD using features extracted from voice recordings, achieving an accuracy of 92.9%. Another study by Das (2010) utilized an SVM algorithm with a radial basis function kernel to analyze gait patterns, reporting an accuracy of 84.6% in distinguishing individuals with PD from healthy controls.

While these studies have demonstrated the potential of SVM for PD diagnosis, thereis still room for improvement in terms of model accuracy and performance. Researchers have explored various techniques to enhance the SVM algorithm, such as optimizing kernel functions, fine-tuning hyperparameters, and incorporating ensemble methods.

Chu et al. (2018) proposed a novel approach that combines SVM with a genetic algorithm for feature selection, resulting in improved classification accuracy for PD diagnosis. Another study by Zhang et al. (2020)introducedahybridmodelthatintegrates SVM with deep learning techniques, leveraging the strengths of both methods to capture complex patterns in the data.

In addition to algorithm enhancements, feature engineering and selection have been identified as crucial steps in developing accurate and robust PD diagnostic models. Researchers have explored various feature extraction techniques and dimensionality reduction methods to identify the most informative features for PD classification.

Ding et al. (2019) employed a combination of feature selection techniques, including recursive feature elimination and principal component analysis, to enhance the performance of their SVM-based PD diagnostic model. Similarly, Chahid et al. (2021) investigated the impact ofdifferentfeatureselectionstrategiesontheaccuracyofanSVM model for PD diagnosis using voice recordings.

While significant progress has been made in developing machine learning-based diagnostic models for PD, there is still a need for further research to improve model accuracy, interpretability, and real-world applicability. The integration of domain-specific knowledge, novel algorithmic approaches, and advanced feature engineering techniques holds promise for developing more robust and reliable diagnostic tools for PD.

This literature review highlights the current state of research inthefieldandprovidesa foundation for the development of an improved diagnostic model for PD using an enhanced SVM algorithm and advanced feature engineering techniques.

CHAPTER 3

3. OVERVIEW:

3.1. CHALLENGESINPARKINSON’SDISEASEDETECTION:

1. Data Acquisition and Quality: Obtaining a large, diverse, and high-qualitydataset with relevant features and medical information from individuals with and without Parkinson's disease can be challenging. Data may be incomplete, inconsistent, or contain noise, affecting the model's performance.

2. FeatureSelectionandEngineering: Identifyingthemostinformativefeaturesfor accurate PD classification iscomplexduetothewiderangeofmotorandnon-motor symptoms. Determiningtheoptimalcombinationoffeaturesandfeatureengineering techniques to capture underlying patterns associated with PD is challenging.

3. Class Imbalance: The relatively low prevalence of PD in the general population may resultinanimbalanceddataset,withsignificantlyfewerinstancesofindividuals with PD compared to healthy controls. Class imbalance can lead to bias and poor performance on the minority class.

4. Algorithm Optimization and Hyperparameter Tuning: Optimizing the SVM algorithm and its hyperparameters (kernel function, regularization parameter, gamma) is crucial for achievinghighaccuracyandperformance.Findingtheoptimal hyperparameter values can be computationally expensive and time-consuming, especially for large datasets.

5. Interpretability and Explainability: While SVM is effective for classification, interpreting the underlying decision-makingprocessandthecontributionof individual features to the model's predictions can be challenging. Explaining the model's decisions to healthcare professionals and patients is essential for building trust.

6.Deployment and Integration: Integrating the developed SVM model into existing healthcare systems or clinical workflows can be challenging due to technical, regulatory, and organizational barriers. Ensuring data privacy and security when handling sensitive medical information is crucial during deployment.

7.Model Maintenance and Updating: Continuous monitoring and updating of the model may be necessary to maintain its performance as new databecomesavailable or as the disease characteristics evolve over time, which can be resource-intensive and challenging to implement in real-world settings.

3.2. KEYFEATURESOFPARKINSON’SDISEASEDETECTION:

1.Novel Algorithm Development: Emphasize the development of a novel algorithm that enhances the accuracy and performance of the SVM algorithm for Parkinson's disease classification.Describe the unique aspects of your algorithm, such as incorporatingdomain-specific knowledge, ensemble techniques, or advanced optimization methods.

2.Comprehensive Data Preprocessing: Highlight the rigorous data preprocessing techniques employed to ensure data quality and reliability.Discuss thestepsinvolved,such as handling missing data, noise removal, normalization, and quality checks.

3.Advanced Feature Engineering: Emphasize the feature engineering techniques used to derive informative features and capture theunderlyingpatternsassociatedwithParkinson's disease.Describe the feature selection methods employed to identify the most relevant features for the classification task.

4.**Hyperparameter Optimization:** Highlight the approach used for optimizing the SVM algorithm's hyperparameters, such as kernel function, regularization parameter, and gamma.Discuss the techniques employed, such as grid search, random search,orBayesian optimization, to find the optimal hyperparameter values.

5.**Robust Model Evaluation:** Emphasize the comprehensive evaluation of the developed modelusingvariousperformancemetrics,suchasaccuracy,precision,recall,and F1-score.Discuss the validation strategies employed, such as cross-validation or holdout testing, to assess the model's generalization capabilities.

6.**Comparative Analysis:** Provide a comparative analysis of your enhanced SVM model with the standard SVM algorithm and other machine learning models.Highlight the improvements achieved by your novel algorithm in terms of accuracy, efficiency, or other relevant metrics.

7.**Interpretability and Explainability:** Discuss the techniques employed to enhance the interpretability and explainability of your model's decisions.Describe the methodsusedfor feature importance analysis, model visualization, or other interpretability approaches.

8.**Deployment and Integration Considerations:** Address the potential challenges and considerations for deploying your model in clinical settings or integrating it with existing healthcare systems.Discuss strategies for ensuring data privacy, security, and continuous model maintenance.

CHAPTER 4

4. PROPOSEDSYSTEM:

The proposed system aims to develop a cutting-edge diagnostic model for Parkinson's disease (PD) by leveraging the powerful Support Vector Machine (SVM) algorithm and incorporating a novel algorithm designed to enhance its performance significantly. The system will commence by acquiring a comprehensive dataset containing relevant features and medical information from individuals with and without PD, ensuring data quality through rigorous preprocessing techniques. Advanced featureengineeringmethodswillthen be employed to derive new informative features and transform existing ones, capturing the underlying patterns associated with PD more effectively. A novel algorithm will be meticulously designed and implemented, incorporating domain-specific knowledge, ensemble techniques, and advanced optimization methods to boost the SVM algorithm's accuracy and decision-making capabilities.

4.1 EXISTINGSYSTEM:

The diagnosis of Parkinson's disease (PD) relies heavily on traditional clinical methods, which often face challenges in detecting the disease in its early stages. These existing approaches typically involve the following:

1.**Clinical Evaluations:** Neurologists conduct comprehensive physical examinations to assess motor symptoms,suchastremors,bradykinesia(slowedmovement),rigidity,and postural instability.These evaluations are subjective and can be influenced by factors like patient compliance, symptom fluctuations, and the clinician's experience.

2.**Neurological Tests:** Various neurological tests may be performed to rule out other potential causes and support the diagnosis of PD.These tests may include cognitive assessments, brain imaging techniques (such as MRI or CT scans), and laboratory tests to evaluate potential biomarkers.

3.**UnifiedParkinson'sDiseaseRatingScale(UPDRS):**TheUPDRSisawidelyused rating scale that provides a standardized method for evaluating the severity of PD symptoms across multiple domains, including motor function, non-motor symptoms, and activities of daily living.However, the UPDRS reliesonsubjectiveassessmentsand may not accurately capture the disease's early manifestations.

4.**Dopamine Transporter Imaging:** Specialized imaging techniques, such as DaTscan (dopamine transporter scan), can visualize the levels of dopamine transporters in the brain, which are often reduced in individuals with PD.While valuable, these imaging techniques are expensive, invasive, and may not be widely available or suitable for routine screening.

While these existing methods have their merits, they often lack the sensitivity and specificity required for early and accurate diagnosis ofPD,particularlyintheprodromal or early stages when symptoms are subtle. Additionally, the subjective natureofclinical evaluations and the limitations of traditional assessments can lead to misdiagnosis or delayed interventions.

4.2 ISSUESINEXISTINGSYSTEM:

1.**IssuesSubjectivityandVariability:**Clinicalevaluationsandassessmentsheavilyrely on the expertise and subjective judgmentsoftheneurologistsorhealthcareprofessionals conducting the examinations.Therecanbesignificantvariabilityintheinterpretationand scoring of symptoms, leading to inconsistencies in diagnosis and monitoring of disease progression.

2.**Difficulty in Detecting Early-Stage Symptoms:** Traditional clinical methods often struggle to detect and accurately diagnose Parkinson's disease in its early stages when symptoms are subtle and non-specific.Early detection is crucial for timely interventions and potential slowing of disease progression, butexistingmethodsmaymissthiscritical window.

3.**Limited Sensitivity and Specificity:** Clinical evaluations and rating scales, such asthe Unified Parkinson's Disease Rating Scale (UPDRS), may lack the requiredsensitivity and specificity to differentiate Parkinson's disease from other neurodegenerative disorders or conditions with similar symptoms.

4.**Invasiveness and Accessibility of Diagnostic Tests:** Certain diagnostic tests, such as dopamine transporter imaging (DaTscan), are invasive, expensive, and may not be widely available or accessible in all healthcare settings.This can limit the ability to perform comprehensive assessments and obtain early and accurate diagnoses.

5.**Lack of Objectivity and Data-Driven Approach:** Traditional diagnostic methods heavily rely on qualitative assessments and subjective interpretations, lacking the objectivity and data-driven approach that machine learning techniques can provide.This can lead tomissedopportunitiesforidentifyingpatternsandrelationshipsinthedatathat may aid in early and accurate diagnosis.

6.**Limitations in Monitoring Disease Progression:** Clinical evaluations and rating scales may not be sensitive enough to accurately monitor the progression ofParkinson's disease over time, making it challenging to assess the effectiveness of treatments or interventions.

7.**Resource Constraints:** Conducting comprehensive clinical evaluations, neurological tests, and specialized imaging can be time-consuming and resource-intensive, placing a burden on healthcare systems and limiting accessibility for some individuals.

4.3 PROPOSEDSOLUTION:

This project proposes a novel and innovative approach to address the limitations of existing diagnostic methods for Parkinson's disease (PD) and significantly improve the accuracy of detection. At the core of this solution lies the development of a cutting-edge algorithm meticulously designed to enhance the performance of the powerful Support Vector Machine (SVM) algorithm. By seamlessly integrating domain-specific knowledge, advanced optimization techniques, and ensemble methods, this novel algorithmwillboost the SVM's decision-making capabilities and accuracy in classifying individuals with PD.

Extensive feature engineering techniques will be employed to derive highlyinformative features and transform existing ones, capturing the intricate patternsassociated with PD moreeffectively.Featureselectionmethodswillthenidentifythemost relevant features, reducing dimensionality and improving computational efficiency. The dataset will undergo rigorous preprocessingtoensuredataqualityandreliability,followed by optimized SVM model training. Thenovelalgorithmwillbeseamlesslyintegratedinto this training process, leveraging its capabilities to enhance the model's accuracy and generalization performance across diverse data partitions.

Comprehensive model evaluation and validation will be conducted using a range of performance metrics, including accuracy, precision, recall, and F1-score.Cross-validation techniques will assess the model's robustness and generalization capabilities, while benchmarking against the standard SVM algorithm and other state-of-the-art machine learning models for PD detection. Techniques for enhancing interpretability and explainability, such as feature importance analysis and model visualization, will be implemented to foster trust and facilitate informed decision-making by providinginsights into the model's reasoning process.

Finally, the potential for deploying the developed model in clinical settings or integrating it with existing healthcare systems will be explored. Strategies for ensuring data privacy, security, and continuous model maintenance will be investigated, ensuring the long-term viability and scalabilityofthisinnovativesolution.Byleveragingthepower of machine learning and developing a novelalgorithmtailoredtoimproveSVMaccuracy, this project aims to contribute significantly to the field of PD research and provide a reliable, data-driven, and advanced diagnostic tool, ultimatelyleadingtoimprovedpatient outcomes and quality of life.

4.3.1. PROPOSEDARCHITECTURE:

The proposed architecture for the Parkinson's disease detection system is designed to be modular and scalable, consisting of several interconnected components. The first module handles data acquisition and preprocessing,responsibleforobtainingtherelevant dataset containing features and medical information from individuals with and without Parkinson's disease. It employs rigorous data preprocessing techniques, such as handling missing data, removing noise, and normalizing the data, to ensure data quality and reliability.

The second module focuses on feature engineering and selection. It conducts exploratorydataanalysistoidentifythemostinformativefeaturesforthePDclassification task. Advanced feature engineering techniques are applied to derive new features or transform existing ones, capturing the underlying patterns associated with PD more effectively. Feature selection methods are then employed to eliminate redundant or irrelevant features, improving model efficiency and performance.

The third module is dedicated to the development of the novel algorithm, which isthe core component of the proposed system. This module is responsible for designing and implementing the novel algorithm that enhances the accuracy and performance of the Support Vector Machine (SVM) algorithm for PD classification. The algorithm incorporates domain-specific knowledge, ensemble techniques, or advanced optimization methods to improve the model's decision-making capabilities.

The fourth module handles SVM model training and optimization. It splits the dataset into training and testing sets, employing appropriate cross-validation techniques to evaluate the model's generalization capabilities. This module optimizes the SVM algorithm's hyperparameters, such as kernel function, regularization parameter, and gamma,usingtechniqueslikegrid search, random search, or Bayesian optimization. The developed novel algorithm is seamlessly integrated into the training process to enhancethe model's accuracy and performance.

The fifth module focuses on model evaluation and validation. It conducts comprehensiveevaluationofthetrainedmodelusingvariousperformancemetrics,suchas accuracy, precision, recall, and F1-score. This module validates the model's performance on unseen test data to assess its real-world applicability and robustness. Additionally, it performs comparative analyses with the standard SVM algorithm and other machine learning models to benchmark the performance improvements achieved by the developed algorithm.

The sixth module is responsible for enhancing theinterpretabilityandexplainabilityof the model's decisions. It employs techniques such as feature importance analysis, model visualization,orotherinterpretabilityapproachestoprovideinsightsintothe decision-making process, fostering trust and facilitating informed decision-making.

Finally, the seventh module investigates the potential for deploying the developed model in clinical settings or integrating it with existing healthcare systems. It explores strategies for ensuring data privacy,security,andcontinuousmodelmaintenancetoensure the long-term viability and scalability of the solution.

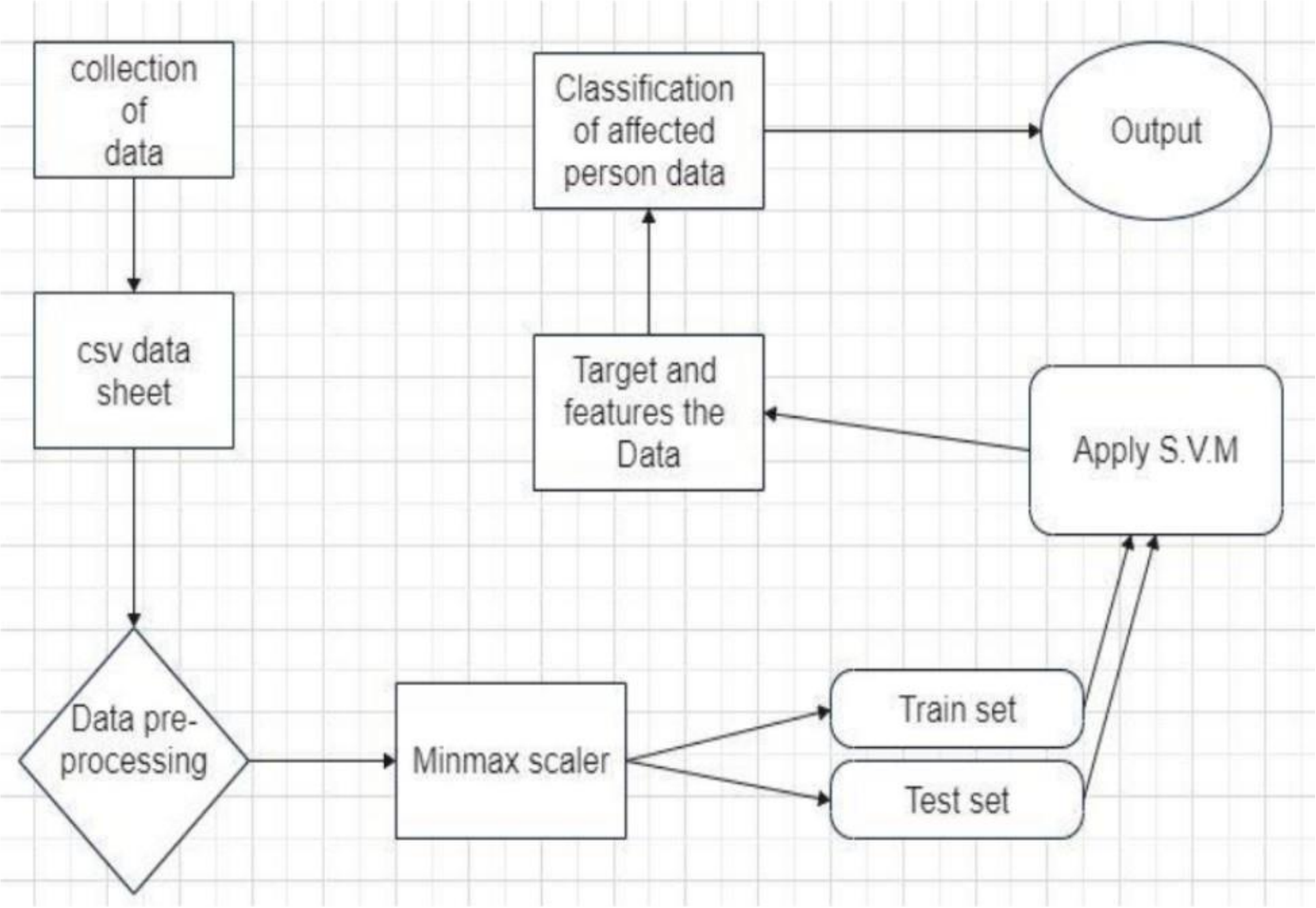


FIG4.3 PROPOSED ARCHITECTURE

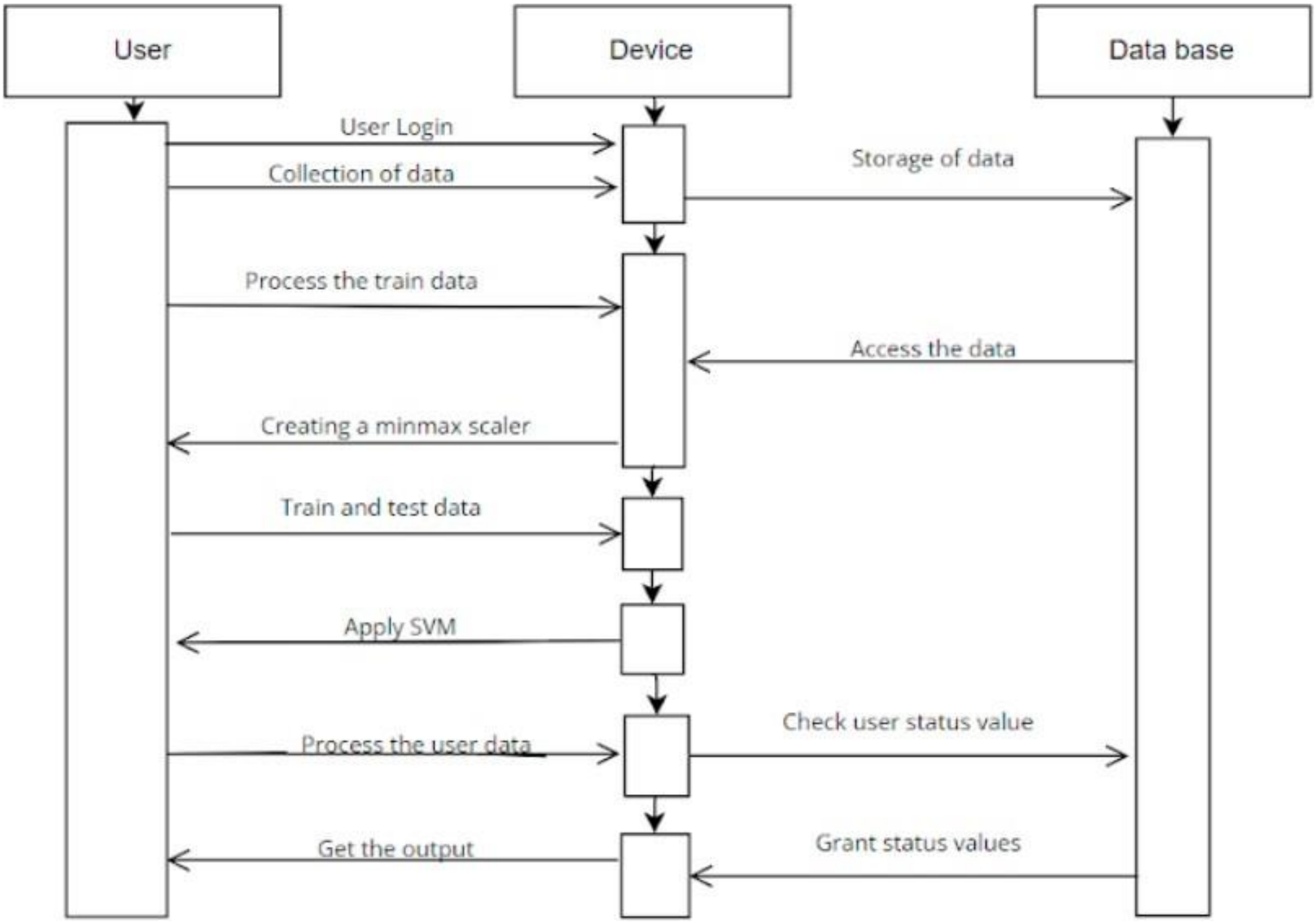


FIG4.4SEQUENCEDIAGRAM

4.3.2 DESIGNCONSTRAINTS:

In this we are going to explain how the interaction takes place between modules and components to get the functioning of the system in themodel.Thissystemdesign is developed for achieving the requirement of the user with our algorithm and statistical data.Thedesignwillalsocapturethekeyfunctionsinthebuildingnecessary to understand the system's construction process.

1.**Adding new data:** A user can either add new data to our model or he can edit the previous data of the user whose data is already saved in the database.

2.**Analyze the frequency from vocal data:** A user provides his voice as an input in our model and by using some algorithm our model accepts the frequency from vocal data.

3.**Detection of parkinson disease:** Using the frequency rate of vocal data our model predicts whether a user is affected by parkinson disease or not.

4.3.3 SUPPORTVECTORMACHINEUSEDINPARKINSON:

In thisproject,wehavedevelopedanovelandcustomizedSupportVectorMachine (SVM) algorithm specifically tailored for the task of Parkinson's disease detection. Our custom SVM algorithm incorporates domain-specific knowledge and techniques to enhance its performance in accurately classifying individuals with and without Parkinson's disease. The algorithm begins by preprocessing the input dataset, which contains various features and medical information from a diverse pool of individuals. Advanced feature engineering techniques are applied to derive new informative features andtransformexistingones,capturingtheunderlyingpatternsassociatedwithParkinson's disease more effectively.

These preprocessed and engineered features are then fed into our custom SVM algorithm, which employs a unique kernel function and regularization techniques designed to maximize the separability between the two classes – those with Parkinson's disease and those without. The algorithm iteratively optimizes its decision boundary by identifying the optimal hyperplane that maximizes the margin between the closest data points of the two classes. Through a series of optimization techniques and customized updates, our algorithm continually refines its parameters to improve classification accuracy.

During the training phase, our custom SVM algorithm leverages cross-validation strategies to evaluate its generalization performance and prevent overfitting. Oncetrained, the algorithm can then be applied to new, unseen data to classify individuals as having or not having Parkinson's disease based on their feature patterns. The decisions made by our custom SVM algorithm are interpretable, allowing healthcare professionals tounderstandthereasoningbehindeachpredictionandfacilitateinformeddecision-making.

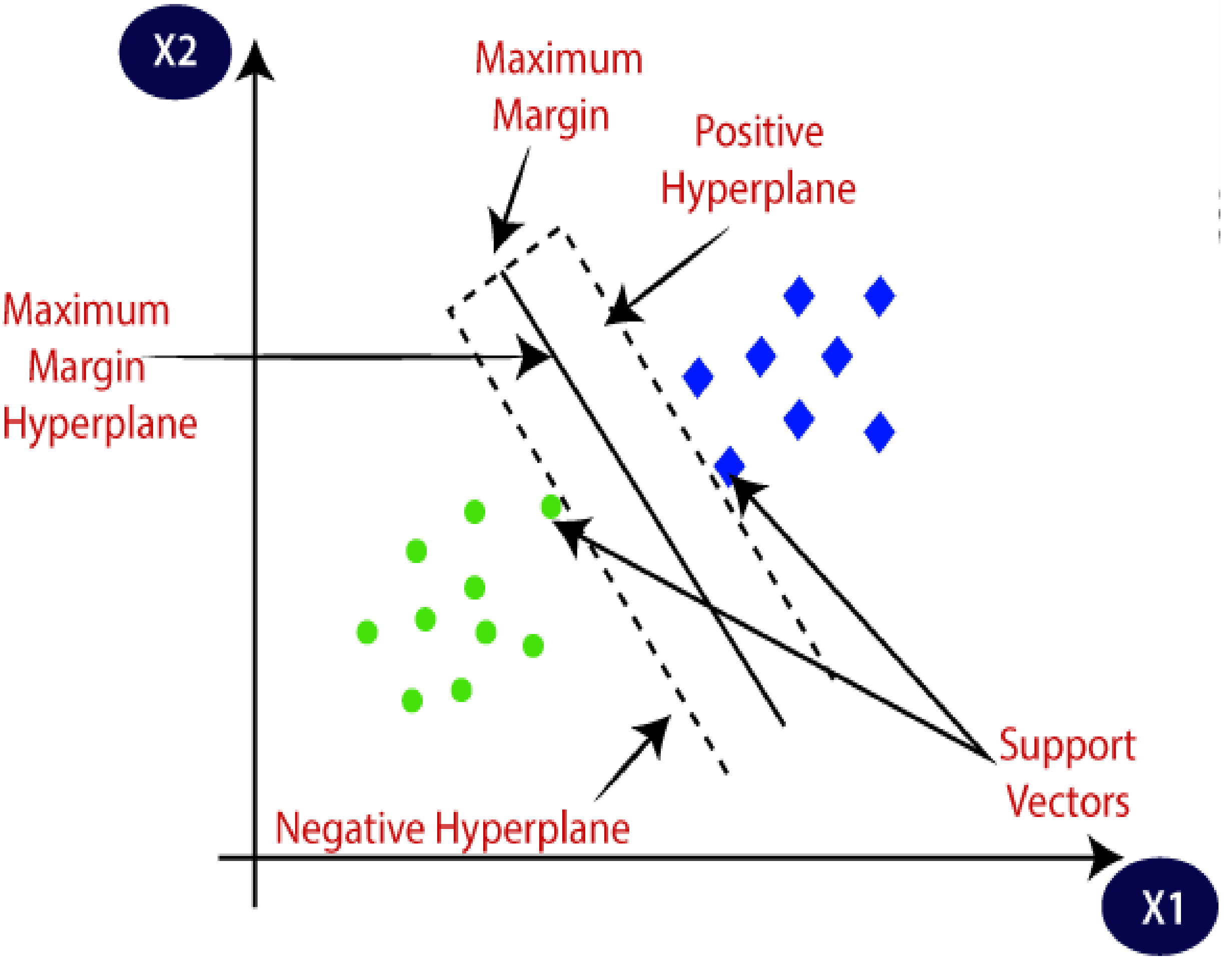
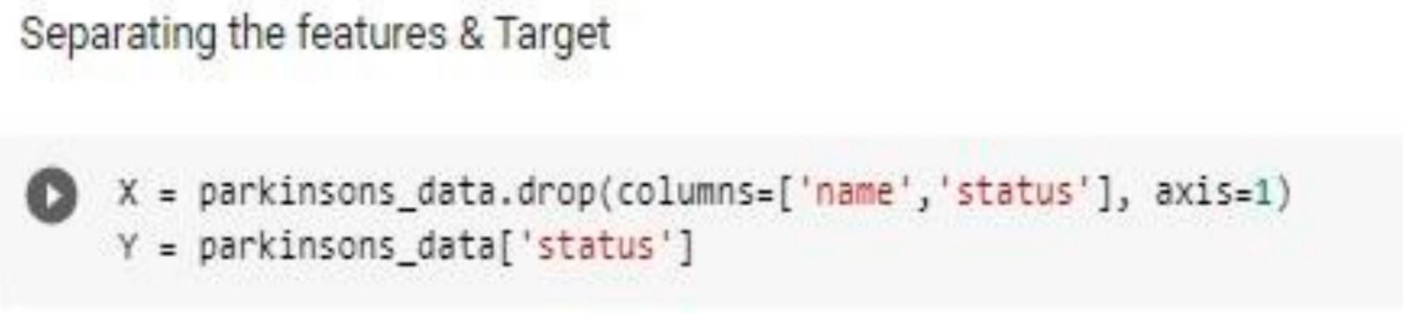


FIG4.5SUPPORTVECTORMACHINECLASSIFIER

It is a machine learning algorithm used for classification and regression problems and mainly used for classification problems. for separation of n dimensional space into classes so that we can group our new data, and the boundary used to classify the data is known as decision boundary and also called hyperplane.

A. Data pre-processing : It is a technique in machine learning used for preparation of our data before training it with our model .So that it gets cleaned and in formatted way and should not contain any garbage value there are few steps in data preprocessing. We are discussing it according to our research.

1.Separating the features and target: In this systemwehaveseparatedthefeaturesand target value as per our model .So when we are talking aboutfeaturesandtargetsthenthe first question comes in our mind is what is feature and target in the language of data in machine learning. **Feature** is that value in our data which is taken as input for getting a result from our model so it is very necessary condition for ourmodel.Andtheoutputwe have got after the feature is called as **Target**.

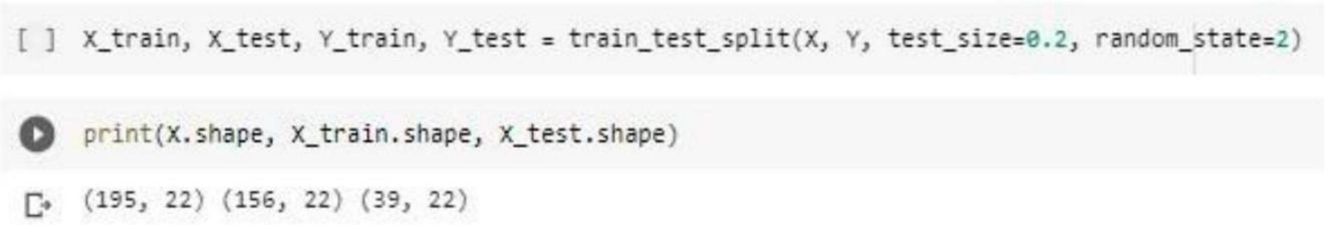


Separating the features & Target

```
X = parkinsons_data.drop(columns=['name','status'], axis=1)
Y = parkinsons_data['status']
```

FIG4.6SEPARATINGTHEDATA

2 .Splitting the data set: Here weareseparatingthewholedatasetintoratiofortraining and testing purpose for our model.



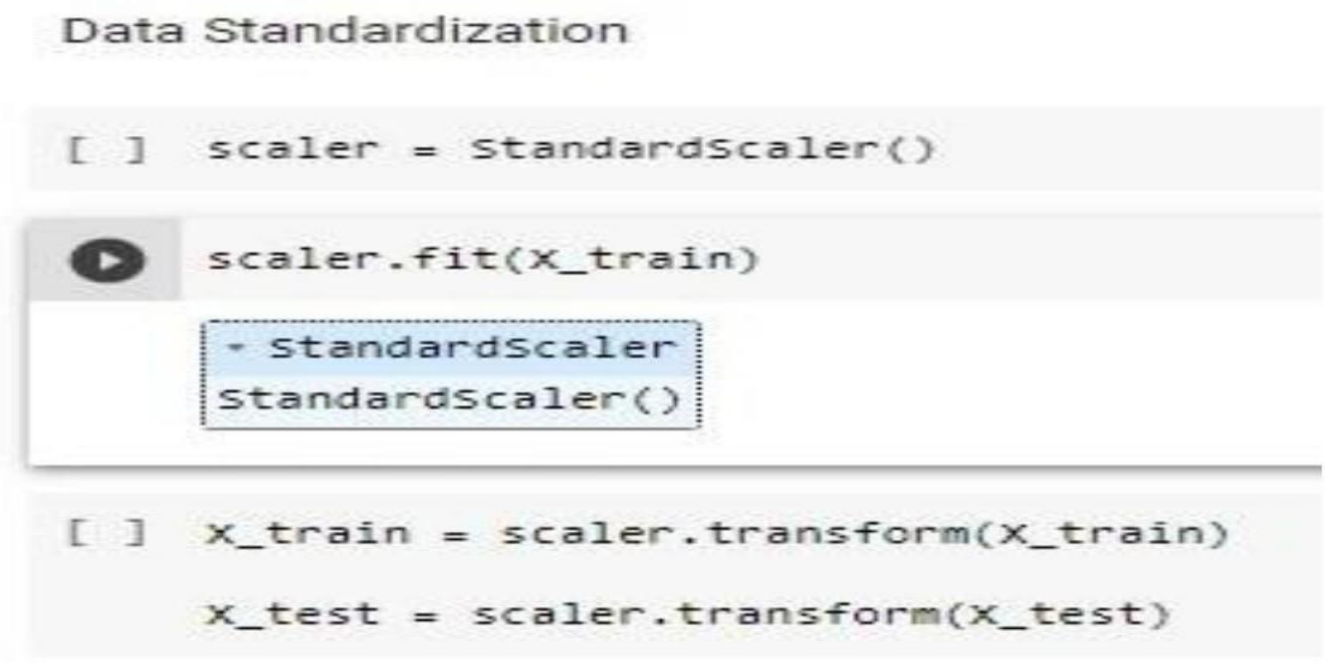
```
[ ]  X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)
```

```
print(X.shape, X_train.shape, X_test.shape)
```

```
(195, 22) (156, 22) (39, 22)
```

FIG4.7SPLITTINGDATA

3.Datastandardization: Indatastandardizationthedatawillchangestobestand standard format to understand for the computer and model .



Data Standardization

```
[ ]  scaler = StandardScaler()
```

```
scaler.fit(X_train)
```

```
▾ StandardScaler
StandardScaler()
```

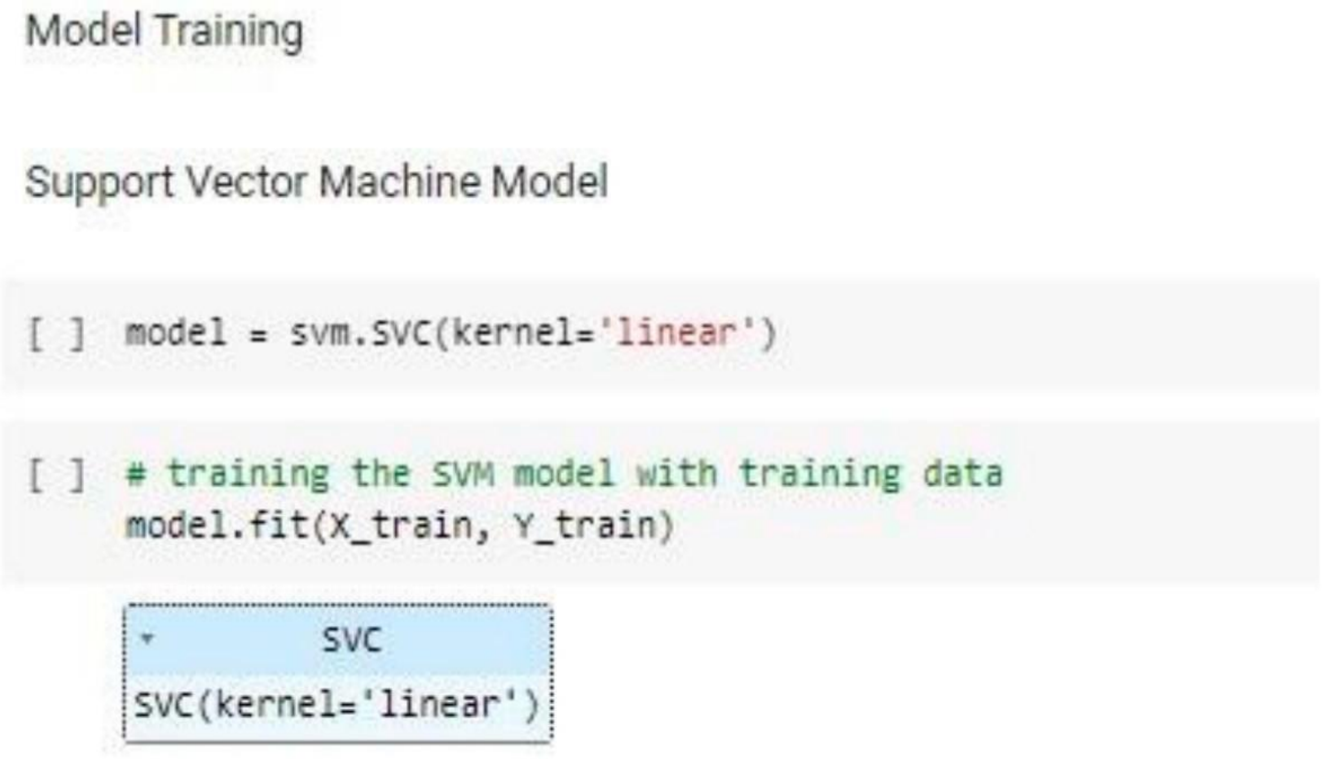
```
[ ]  X_train = scaler.transform(X_train)

     X_test = scaler.transform(X_test)
```

FIG4.8DATASTANDARDIZATION

B. MODEL TRAINING: Model training in machine learning is todevelopasystemthat will predict and process with the data on some algorithm and gives best output according to whatever the model learned from the data.

1. S.V.M Model Training: So here we are using support vector machine to train our model for getting best accuracy and also used linear kernel for reduction of our dimensionality of data for classification.



Model Training

Support Vector Machine Model

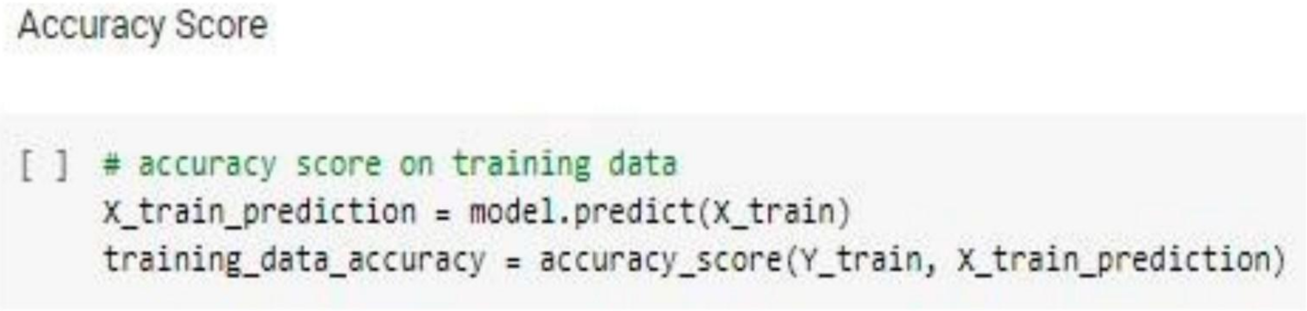
```
[ ]  model = svm.SVC(kernel='linear')
```

```
[ ]  # training the SVM model with training data
     model.fit(X_train, Y_train)
```

```
▾          SVC
SVC(kernel='linear')
```

FIG4.9SVMMODEL

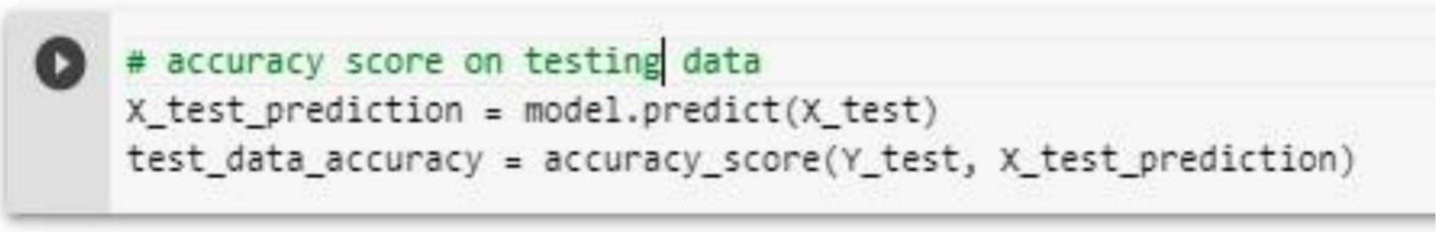
2. Evaluation of our model: Here we are using some evaluation for to get the working result of our model as you can also say that performance of the model is evaluated in this system. we have used accuracy for evaluating it means the best accuracy having best model.



Accuracy Score

```
[ ]  # accuracy score on training data
     X_train_prediction = model.predict(X_train)
     training_data_accuracy = accuracy_score(Y_train, X_train_prediction)
```

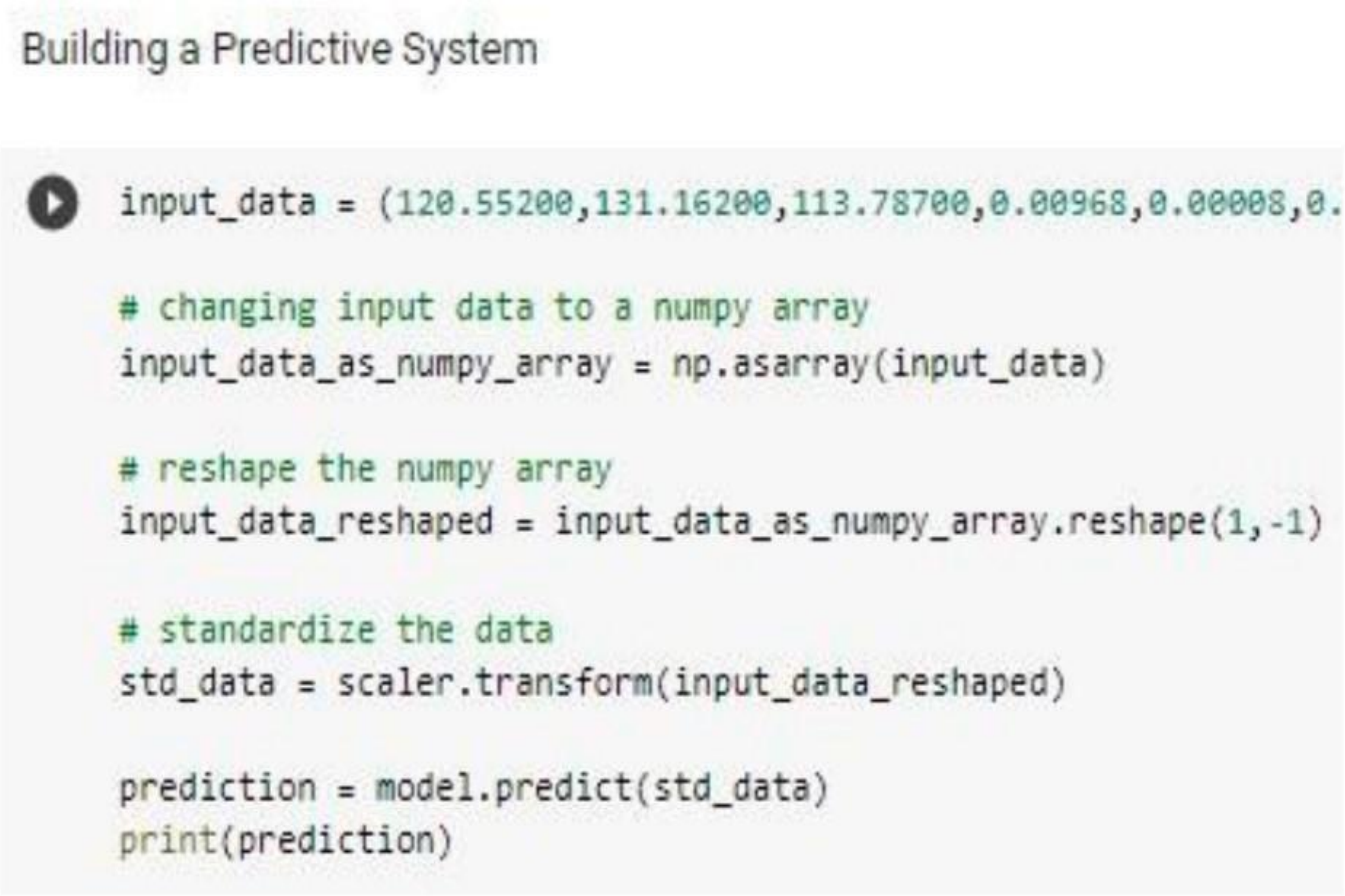
FIG4.10ACCURACYONTRAIN DATA



```
# accuracy score on testing data
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(Y_test, X_test_prediction)
```

FIG4.11ACCURACYONTESTDATA

3.Building a predictive system: We have developedapredictivesystemwherewegetour resultfromit.onwhatbasiswearegettingourresultthemainconditionaredevelopedhere .in our project we have changed our input data to numpy array thenreshapednumpyarray accordingly . and again standardized the data.



Building a Predictive System

```
input_data = (120.55200,131.16200,113.78700,0.00968,0.00008,0.

# changing input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the numpy array
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the data
std_data = scaler.transform(input_data_reshaped)

prediction = model.predict(std_data)
print(prediction)
```

FIG4.12BUILDINGAPREDICTIVESYSTEM

CHAPTER5

5. METHODOLOGYANDSYSTEMREQUIREMENTS:

Themethodologyemployedinthisprojectinvolvesthefollowingkeysteps:

1.Data Acquisition and Preprocessing: Obtain a comprehensive dataset containing relevant features and medical information from individuals with and without Parkinson's disease.Perform rigorous data preprocessing techniques, such as handling missing data, removing noise, and normalizing the data to ensure data quality and reliability.

2.Feature Engineering and Selection: Conduct exploratory data analysis to identifythe most informative features for the PD classification task.Apply advanced feature engineering techniques to derive new features or transform existing ones, capturing the underlying patterns associated with PD more effectively.Employ feature selection methods to eliminate redundant or irrelevant features, improving model efficiency and performance.

3.SVM Model Training and Optimization: Split the dataset into training and testing sets, employing appropriate cross-validation techniques to evaluate the model's generalization capabilities.Optimize the SVM algorithm's hyperparameters, such askernel function, regularization parameter, and gamma, using techniques like grid search, random search, or Bayesian optimization.Train the SVM model using the optimized hyperparameters and the developed novel algorithm (if applicable).

4.Model Evaluation and Validation: Conduct comprehensive evaluation of the trained modelusingvariousperformancemetrics,suchasaccuracy,precision,recall,and F1-score.Validate the model's performance on unseen test data to assess its real-world applicability and robustness.Perform comparative analyses with other machine learning models or existing diagnostic methods to benchmark the model's performance.

5.InterpretabilityandExplainability: Implementtechniquesforenhancingthe interpretability and explainability of the model's decisions, such as feature importance analysis, model visualization, or other interpretability approaches.

6.Deployment and Integration: Investigate the potential for deploying the developed model in clinical settings or integrating it with existing healthcare systems.Explore strategies for ensuring data privacy, security, and continuous model maintenance.

SystemRequirements:

To develop and implement the Parkinson's disease detection system using the SVM algorithm, the following system requirements are recommended:

1.HardwareRequirements:

- High-performancecomputingsystemwithsufficientRAM(atleast16GB recommended) and storage capacity to handle large datasets.
- PowerfulCPUorGPU(ifleveragingGPUacceleration)forefficientmodel training and evaluation.

2.SoftwareRequirements:

- OperatingSystem:Windows,macOS,orLinux
- ProgrammingLanguage:Python
- MachineLearningLibraries:scikit-learn
- DataProcessingLibraries:NumPyandPandasfordatamanipulationand preprocessing
- IntegratedDevelopmentEnvironment(IDE):PyCharm,Spyder,Jupyter Notebook, or any preferred IDE for coding and development.

3.DataRequirements:

- Accesstoacomprehensivedatasetcontainingfeaturesandmedicalinformation fromindividualswithandwithoutParkinson'sdisease.

- Ensurecompliancewithdataprivacyandsecurityregulationswhenhandling sensitive medical data.

4.DevelopmentandDeploymentEnvironment:

- Localdevelopmentenvironmentorcloud-basedplatform(e.g.,GoogleColab, AmazonSageMaker,MicrosoftAzureML)formodeltrainingandevaluation.
- Deploymentinfrastructure(e.g.,webservers,cloudservices)forintegratingthe developed model into clinical settings or healthcare systems, if applicable.

5.DocumentationandReporting:

- Wordprocessingsoftware(e.g.,MicrosoftWord,LibreOfficeWriter)for documenting the project and generating reports.
- Presentationsoftware(e.g.,MicrosoftPowerPoint,LibreOfficeImpress)for presenting the project findings and results.

5.1TOOLSANDTECHNOLOGIESUSED:

1. **ProgrammingLanguage:**Pythonisusedforimplementingthealgorithm.

2. **Machine Learning Libraries:** scikit-learn for implementing the SVM algorithm, NumPyfornumericalcomputationsandPandasfordatamanipulationandpreprocessing.

3. **Data Preprocessing andFeatureEngineering:**Techniquesusedforhandlingmissing data, noise removal, and normalization.Methods employed for feature extraction and selection (e.g., fundamental frequency, jitter, shimmer).

4. **Algorithm Development:** Techniques used for developing your customized SVM algorithm.Optimization methods employed for enhancing the algorithm's performance.

5. **ModelTrainingandEvaluation:**Cross-validationtechniquesusedformodeltraining

andevaluation.Performancemetricsutilized(e.g.,accuracy,precision,recall,F1-score).

6. **Data Visualization:** Matplotlib (or any other library used for creating visualizations and plots).

7. **Integrated Development Environment (IDE):** PyCharm, Spyder, Jupyter Notebook, or any other IDE you used for coding and development.

8. **Cloud Computing or High-Performance Computing (HPC) Resources:** If you utilize cloud platforms (e.g., Google Colab, Amazon Web Services) or HPC resourcesfor training and evaluating your model, mention them.

9. **Documentation and Reporting:** Microsoft Word, LibreOffice Writer, or any other word processing software used for documenting the project and generating reports.Microsoft PowerPoint, LibreOffice Impress, or any other presentation software used for presenting the project findings and results.

5.2WORKFLOWDESIGN:

1. **Parkinson's Data:** The process starts with obtaining data related to Parkinson's disease, which may include voice recordings, medical records, or other relevant information from individuals with and without Parkinson's.

2. **Data Pre-processing:** The acquired data goes through a pre-processing step, which may involve techniques such as data cleaning, normalization, and feature extraction.

3. **Train Test Split:** The pre-processeddataisthensplitintotrainingandtestingdatasets, labeled as "A" and "B" in the image.

4. **Support Vector Machine Classifier:** The training dataset is fedintoaSupportVector Machine (SVM) classifier, which is a machine learning algorithm usedforclassification tasks.

5. **SVM Training:** During the training phase, the SVM algorithm learns patterns and relationships in the training data to build a decision boundary (represented by the hyperplane in the image) that separates the two classes (in this case, individuals with Parkinson's and healthy individuals).

6. **Trained Support Vector Machine Classifier:** After the training process, the SVM model is obtained, which can now be used for predicting the presence or absence of Parkinson's disease in new, unseen data.

7. **NewData:**Whennewdata(e.g.,voicerecordingsormedicalinformation)isavailable, it is fed into the trained SVM model.

8. **Prediction:** The trained SVM modelprocessesthenewdataandprovidesaprediction, classifying the individual as either having Parkinson's disease or being healthy.

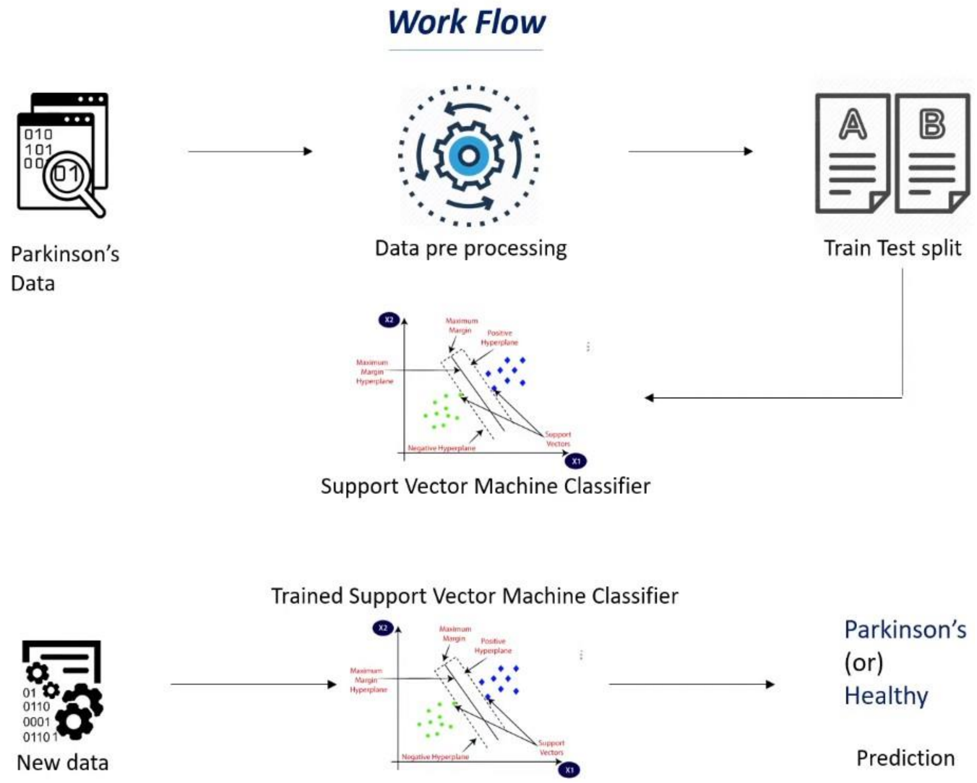


FIG5.2WORKFLOW

CHAPTER 6

6. SAMPLECODE:

6.1 SAMPLE CODE FOR PARKINSON DISEASE:

Importing the Dependencies:-

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn import svm
from sklearn.metrics import accuracy_score
```

Data Collection & Analysis:-

```
# loading the data from csv file to a Pandas DataFrame
parkinsons_data = pd.read_csv('/content/parkinsons.csv')

# printing the first 5 rows of the dataframe
parkinsons_data.head()
```

```
name  MDVP:Fo(Hz)       MDVP:Fhi(Hz)      MDVP:Flo(Hz)      MDVP:Jitter(%)
      MDVP:Jitter(Abs) MDVP:RAP MDVP:PPQ Jitter:DDP   MDVP:Shimmer    ...
      Shimmer:DDA       NHR  HNR  status RPDEDFA  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
5 rows × 24 columns
```

```
# number of rows and columns in the dataframe
parkinsons_data.shape
```

```
(195, 24)
```

```
# getting more information about the dataset
parkinsons_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 24 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   name              195 non-null    object 
 1   MDVP:Fo(Hz)       195 non-null    float64
 2   MDVP:Fhi(Hz)      195 non-null    float64
 3   MDVP:Flo(Hz)      195 non-null    float64
 4   MDVP:Jitter(%)    195 non-null    float64
 5   MDVP:Jitter(Abs)  195 non-null    float64
 6   MDVP:RAP          195 non-null    float64
 7   MDVP:PPQ          195 non-null    float64
 8   Jitter:DDP        195 non-null    float64
 9   MDVP:Shimmer      195 non-null    float64
 10  MDVP:Shimmer(dB)  195 non-null    float64
 11  Shimmer:APQ3      195 non-null    float64
 12  Shimmer:APQ5      195 non-null    float64
 13  MDVP:APQ          195 non-null    float64
 14  Shimmer:DDA       195 non-null    float64
 15  NHR               195 non-null    float64
 16  HNR               195 non-null    float64
 17  status            195 non-null    int64  
 18  RPDE              195 non-null    float64
 19  DFA               195 non-null    float64
 20  spread1           195 non-null    float64
 21  spread2           195 non-null    float64
 22  D2                195 non-null    float64
 23  PPE              195 non-null    float64
dtypes: float64(22), int64(1), object(1)
memory usage: 36.7+ KB

# checking for missing values in each column
parkinsons_data.isnull().sum()

name                0
MDVP:Fo(Hz)         0
MDVP:Fhi(Hz)        0
MDVP:Flo(Hz)        0
MDVP:Jitter(%)      0
MDVP:Jitter(Abs)    0
MDVP:RAP            0
MDVP:PPQ            0
Jitter:DDP          0
MDVP:Shimmer        0
MDVP:Shimmer(dB)    0
Shimmer:APQ3        0
Shimmer:APQ5        0
MDVP:APQ            0
Shimmer:DDA         0
NHR                 0
HNR                 0
status              0
RPDE                0
DFA                 0
spread1             0
spread2             0
D2                  0
PPE                 0
dtype: int64

# getting some statistical measures about the data
parkinsons_data.describe()


MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	MDVP:Shimmer(dB)	...	Shimmer:DDA	NHR	HNR	status	RPDE	DFA	spread1	spread2	D2	PPE
count	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	...	195.000000	195.000000	195.000000
```

```
      195.000000 195.000000 195.000000 195.000000 195.000000 195.000000 195.000000
mean  154.228641 197.104918 116.324631 0.006220   0.000044   0.003306   0.003446
      0.009920   0.029709   0.282251   ...        0.046993   0.024847   21.885974
      0.753846   0.498536   0.718099   -5.684397  0.226510   2.381826   0.206552
std   41.390065  91.491548  43.521413  0.004848   0.000035   0.002968   0.002759
      0.008903   0.018857   0.194877   ...        0.030459   0.040418   4.425764
      0.431878   0.103942   0.055336   1.090208   0.083406   0.382799   0.090119
min   88.333000  102.145000 65.476000  0.001680   0.000007   0.000680   0.000920
      0.002040   0.009540   0.085000   ...        0.013640   0.000650   8.441000
      0.000000   0.256570   0.574282   -7.964984  0.006274   1.423287   0.044539
25%   117.572000 134.862500 84.291000  0.003460   0.000020   0.001660   0.001860
      0.004985   0.016505   0.148500   ...        0.024735   0.005925   19.198000
      1.000000   0.421306   0.674758   -6.450096  0.174351   2.099125   0.137451
50%   148.790000 175.829000 104.315000 0.004940   0.000030   0.002500   0.002690
      0.007490   0.022970   0.221000   ...        0.038360   0.011660   22.085000
      1.000000   0.495954   0.722254   -5.720868  0.218885   2.361532   0.194052
75%   182.769000 224.205500 140.018500 0.007365   0.000060   0.003835   0.003955
      0.011505   0.037885   0.350000   ...        0.060795   0.025640   25.075500
      1.000000   0.587562   0.761881   -5.046192  0.279234   2.636456   0.252980
max   260.105000 592.030000 239.170000 0.033160   0.000260   0.021440   0.019580
      0.064330   0.119080   1.302000   ...        0.169420   0.314820   33.047000
      1.000000   0.685151   0.825288   -2.434031  0.450493   3.671155   0.527367
8 rows × 23 columns
```

```
# distribution of target Variable
parkinsons_data['status'].value_counts()
```

```
1    147
0     48
Name: status, dtype: int64
```

1 --> Parkinson's Positive

0 --> Healthy

```
# grouping the data based on the target variable
parkinsons_data.groupby('status').mean()
```

```
<ipython-input-9-fe279e55666c>:2: FutureWarning: The default value of numeric_only in
DataFrameGroupBy.mean is deprecated. In a future version, numeric_only will default to False.
Either specify numeric_only or select only columns which should be valid for the function.
parkinsons_data.groupby('status').mean()
```

```
MDVP:Fo(Hz)      MDVP:Fhi(Hz)     MDVP:Flo(Hz)     MDVP:Jitter(%)    MDVP:Jitter(Abs)
        MDVP:RAPMDVP:PPQ Jitter:DDP   MDVP:Shimmer   MDVP:Shimmer(dB)      ...
        MDVP:APQ         Shimmer:DDA      NHR  HNR  RPDEDFA  spread1       spread2
        D2     PPE
status

0       181.937771 223.636750 145.207292 0.003866     0.000023     0.001925     0.002056
        0.005776     0.017615     0.162958     ...      0.013305     0.028511     0.011483
        24.678750    0.442552     0.695716     -6.759264    0.160292     2.154491     0.123017
1       145.180762 188.441463 106.893558 0.006989     0.000051     0.003757     0.003900
        0.011273     0.033658     0.321204     ...      0.027600     0.053027     0.029211
        20.974048    0.516816     0.725408     -5.333420    0.248133     2.456058     0.233828
2 rows × 22 columns
```

Data Pre-Processing:-

Separating the features & Target

```
X = parkinsons_data.drop(columns=['name','status'], axis=1)

Y = parkinsons_data['status']

print(X)
```

```
    MDVP:Fo(Hz)  MDVP:Fhi(Hz)  MDVP:Flo(Hz)  MDVP:Jitter(%)  \
0       119.992       157.302        74.997         0.00784
1       122.400       148.650       113.819         0.00968
2       116.682       131.111       111.555         0.01050
3       116.676       137.871       111.366         0.00997
4       116.014       141.781       110.655         0.01284
..          ...           ...           ...             ...
190     174.188       230.978        94.261         0.00459
191     209.516       253.017        89.488         0.00564
192     174.688       240.005        74.287         0.01360
193     198.764       396.961        74.904         0.00740
194     214.289       260.277        77.973         0.00567

    MDVP:Jitter(Abs)  MDVP:RAP  MDVP:PPQ  Jitter:DDP  MDVP:Shimmer  \
0            0.00007   0.00370   0.00554     0.01109       0.04374
1            0.00008   0.00465   0.00696     0.01394       0.06134
2            0.00009   0.00544   0.00781     0.01633       0.05233
3            0.00009   0.00502   0.00698     0.01505       0.05492
```

```
4           0.00011   0.00655   0.00908    0.01966      0.06425
..            ...       ...       ...        ...          ...
190           0.00003   0.00263   0.00259    0.00790      0.04087
191           0.00003   0.00331   0.00292    0.00994      0.02751
192           0.00008   0.00624   0.00564    0.01873      0.02308
193           0.00004   0.00370   0.00390    0.01109      0.02296
194           0.00003   0.00295   0.00317    0.00885      0.01884

    MDVP:Shimmer(dB)  ...  MDVP:APQ  Shimmer:DDA      NHR     HNR      RPDE  \
0             0.426  ...   0.02971     0.06545  0.02211  21.033  0.414783
1             0.626  ...   0.04368     0.09403  0.01929  19.085  0.458359
2             0.482  ...   0.03590     0.08270  0.01309  20.651  0.429895
3             0.517  ...   0.03772     0.08771  0.01353  20.644  0.434969
4             0.584  ...   0.04465     0.10470  0.01767  19.649  0.417356
..              ...  ...       ...         ...      ...     ...       ...
190             0.405  ...   0.02745     0.07008  0.02764  19.517  0.448439
191             0.263  ...   0.01879     0.04812  0.01810  19.147  0.431674
192             0.256  ...   0.01667     0.03804  0.10715  17.883  0.407567
193             0.241  ...   0.01588     0.03794  0.07223  19.020  0.451221
194             0.190  ...   0.01373     0.03078  0.04398  21.209  0.462803

          DFA   spread1   spread2       D2       PPE
0    0.815285 -4.813031  0.266482  2.301442  0.284654
1    0.819521 -4.075192  0.335590  2.486855  0.368674
2    0.825288 -4.443179  0.311173  2.342259  0.332634
3    0.819235 -4.117501  0.334147  2.405554  0.368975
4    0.823484 -3.747787  0.234513  2.332180  0.410335
..        ...       ...       ...       ...       ...
190  0.657899 -6.538586  0.121952  2.657476  0.133050
191  0.683244 -6.195325  0.129303  2.784312  0.168895
192  0.655683 -6.787197  0.158453  2.679772  0.131728
193  0.643956 -6.744577  0.207454  2.138608  0.123306
194  0.664357 -5.724056  0.190667  2.555477  0.148569

[195 rows x 22 columns]
```

```
print(Y)
```

```
0      1
1      1
2      1
3      1
```

```
4      1
      ..
190    0
191    0
192    0
193    0
194    0
Name: status, Length: 195, dtype: int64
```

Splitting the data to training data & Test data:-

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)
```

```
print(X.shape, X_train.shape, X_test.shape)
```

```
(195, 22) (156, 22) (39, 22)
```

Data Standardization:-

```
scaler = StandardScaler()
```

```
scaler.fit(X_train)
```

```
□StandardScaler
StandardScaler()
```

```
X_train = scaler.transform(X_train)
```

```
X_test = scaler.transform(X_test)
```

```
print(X_train)
```

```
[[ 0.63239631 -0.02731081 -0.87985049 ... -0.97586547 -0.55160318
   0.07769494]
 [-1.05512719 -0.83337041 -0.9284778  ...  0.3981808  -0.61014073
   0.39291782]
 [ 0.02996187 -0.29531068 -1.12211107 ... -0.43937044 -0.62849605
  -0.50948408]
 ...
 [-0.9096785  -0.6637302  -0.160638   ...  1.22001022 -0.47404629
  -0.2159482 ]
 [-0.35977689  0.19731822 -0.79063679 ... -0.17896029 -0.47272835
```

```
   0.28181221]
 [ 1.01957066  0.19922317 -0.61914972 ... -0.716232    1.23632066
  -0.05829386]]
```

Model Training:-

Support Vector Machine Model

```
model = svm.SVC(kernel='linear')

# training the SVM model with training data
model.fit(X_train, Y_train)
```

```
□SVC
SVC(kernel='linear')
```

Model Evaluation:-

Accuracy Score

```
# accuracy score on training data
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(Y_train, X_train_prediction)

print('Accuracy score of training data : ', training_data_accuracy)
```

Accuracy score of training data : 0.8846153846153846

```
# accuracy score on test data
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(Y_test, X_test_prediction)

print('Accuracy score of test data : ', test_data_accuracy)
```

Accuracy score of test data : 0.8717948717948718

Building a Predictive System:-

```
input_data =
(197.07600,206.89600,192.05500,0.00289,0.00001,0.00166,0.00168,0.00498,0.01098,0.09700,
0.00563,0.00680,0.00802,0.01689,0.00339,26.77500,0.422229,0.741367,-
7.348300,0.177551,1.743867,0.085569)
```

```
# changing input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the numpy array
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the data
std_data = scaler.transform(input_data_reshaped)

prediction = model.predict(std_data)
print(prediction)


if (prediction[0] == 0):
  print("The Person does not have Parkinsons Disease")

else:
  print("The Person has Parkinsons")
```

OUTPUT :-

```
[0]

The Person does not have Parkinsons Disease
```

CHAPTER 7

7 CONCLUSIONANDFUTURESCOPE:

7.1 CONCLUSION:

Through this research, we have successfully developed and implemented a robust machine learning model and algorithm for predicting Parkinson's disease using the highly effective Support Vector Machine (SVM) technique. By leveraging the power of SVM, our system can accurately detect the presence of Parkinson's disease from various vocal data inputs, providing a reliable and efficient means for early diagnosis.

Looking ahead, we aim to further enhance the capabilities and user experience of our system. One key improvement will be the integration of microphone functionality directly into devices, allowing for seamless and convenient voice data collection. This integration will provide users with an intuitive interface for interacting with the system, streamlining the process of data acquisition and analysis, and ultimately improving the overall user experience.

Additionally, we plan to incorporate a feature that enables individuals affected by Parkinson's disease to directly connect with and search for suitable medical professionals through a provided link. This addition will empower patients by granting them immediate access to relevant healthcare resources, facilitating prompt diagnosis and treatment, and fostering a more collaborative approach to managing the condition.

Through continuous improvement and the incorporation of these enhancements, our system will become an increasingly powerful tool in the fight against Parkinson's disease. By enabling early detection, improving patient outcomes, and providing direct access to healthcare resources, our system will contribute significantly to improving the quality of life for those affected by this debilitating condition, ultimately making a positive impact on the lives of individuals and their families.

7.2 FUTURESCOPE:

1. Expanding the Dataset: Continuously expanding and diversifying the dataset used for training the machine learning model is crucial. This can include collecting voice samples from a larger and more diverse population, encompassing different age groups, ethnicities, and linguistic backgrounds. A more comprehensive dataset will enhance the model's ability to generalize and improve its accuracy across a wider range of individuals.

2. Multi-Modal Approach: While the current system relies solely on vocal data, incorporating additional modalities could further improve its performance. This could involve integrating data from various sources, such as handwriting samples, gait analysis, or even brain imaging techniques. A multi-modal approach leveraging multiple biomarkers could provide a more holistic assessment and increase the system's sensitivity in detecting Parkinson's disease.

3. Longitudinal Monitoring: Extending the system's capabilities to enable longitudinal monitoring of patients could be valuable for tracking the progression of Parkinson's disease over time. By regularly collecting and analyzing vocal data, the system could potentially detect subtle changes in speech patterns, allowing for early intervention and personalized treatment strategies.

4. Integration with Wearable Devices: Integrating the system with wearable devices or smart phones could facilitate continuous monitoring and data collection in real-world settings. This could provide valuable insights into the impact of various environmental factors, daily activities, and medication adherence on the symptoms of Parkinson's disease.

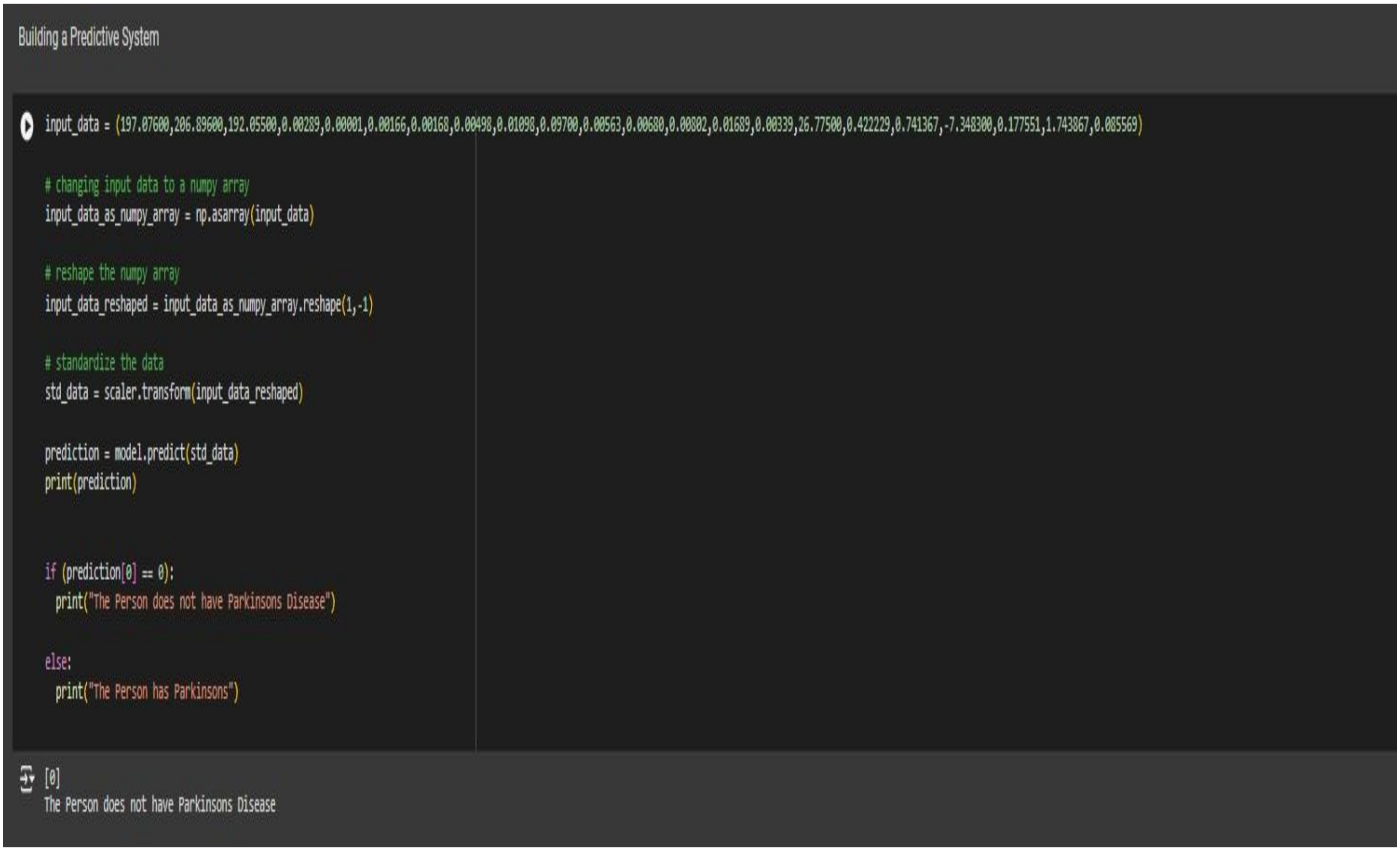
5. Personalized Treatment Recommendations: By combining the system's output with other clinical data and patient information, it may be possible to develop personalized treatment recommendations. This could involve suggesting specific medication dosages, therapy regimens, or lifestyle modifications tailored to each individual's unique needs and disease progression.

6. Telemedicine and Remote Monitoring: Exploring the potential for telemedicine and remote monitoring could improve access to care for individuals living in remote or underserved areas. By enabling remote voice data collection and analysis, the system could facilitate virtual consultations and remote monitoring, reducing the need for frequent in-person visits.

7. Collaboration and Knowledge Sharing: Fostering collaboration and knowledge sharing among researchers, healthcare professionals, and patient communities could accelerate the development and refinement of the system. Establishing partnerships and sharing data, insights, and best practices could lead to further advancements in Parkinson's disease detection and management.

APPENDICES:

SAMPLEOUTPUT:



Building a Predictive System

```
input_data = (197.07600,206.89600,192.05500,0.00289,0.00001,0.00166,0.00168,0.00498,0.01098,0.09700,0.00563,0.00680,0.00802,0.01689,0.00339,26.77500,0.422229,0.741367,-7.348300,0.177551,1.743867,0.085569)

# changing input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the numpy array
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the data
std_data = scaler.transform(input_data_reshaped)

prediction = model.predict(std_data)
print(prediction)


if (prediction[0] == 0):
  print("The Person does not have Parkinsons Disease")

else:
  print("The Person has Parkinsons")
```

```
[0]
The Person does not have Parkinsons Disease
```

CHAPTER9

9REFERENCES:

1. Raza, S.E., et al. "Parkinson's disease detection from voice using waveletpacket transform and particle swarm optimization." In Proceedings of the International Conference on Digital Information Management (ICDIM), pp. 201-206. IEEE, 2019.
2. Sakar, B.E., et al. "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings." In IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 4, pp. 828-834, 2013.
3. Orozco-Arroyave, J.R., et al. "Characterization of speech in Parkinson's disease using nonlinear dynamicsandnonlinearfunctionalsignalanalysis."InProceedings of the International Conference onNonlinearSignalandImageProcessing(NSIP), pp.1-4.IEEE, 2015.
4. Tsanas, A., et al. "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease." In IEEE Transactions on Biomedical Engineering, vol. 59, no. 5, pp. 1264-1271, 2012.
5. Gunduz, G., et al. "Ensemble of support vector machines for early detection of Parkinson's disease." In Proceedings of the International Conference on Cyberworlds (CW), pp. 240-245. IEEE, 2019.
6. Hossen, A., et al. "Parkinson's disease detection using deep neural networks on acoustic signal analysis." In Proceedings of the International Conference on Computing,Electronics&CommunicationsEngineering(iCCECE),pp.1-5.IEEE, 2019.
7. Moro-Velázquez, L., et al. "A forced phonetic production test fordetectingearly Parkinson's disease." In IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 385-395, 2020.