Parkinson's Disease Detection

A MINI PROJECT REPORT

18CSC305J - ARTIFICIAL INTELLIGENCE

Submitted by

CH.TIRU CHANDER REDDY[RA2111026010378]

A.SACHIN VARMA[RA2111026010403]

Under the guidance of

Dr Karpagam M

Assistant Professor, Department of Computer Science and Engineering

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY



S.R.M. Nagar, Kattankulathur, Chengalpattu District

MAY 2024

# SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that Mini project report titled “Parkinson's Disease Detection using Machine

Learning” is the bona fide work of CH TIRU CHANDER REDDY [RA2111026010378] ,SACHINN VARMA [RA2111026010403],who carried out the minor project under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

## SIGNATURE

Dr.Karpagam M

Assistant Professor

CINTEL Faculty of Engineering and Technology

Kattankulathur- Chennai.

ABSTRACT

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that affects a significant portion of the population. Early detection and monitoring of PD symptoms are crucial for effective treatment. In this study, we utilize a dataset comprising biomedical voice measurements from 31 individuals, including 23 with PD. Each row in the dataset represents a voice recording, with columns corresponding to specific voice measures. The primary objective is to discriminate between healthy individuals (status=0) and those with PD (status=1) based on these voice measurements.

We preprocess the dataset, extract relevant features, and employ machine learning techniques, specifically Support Vector Machines (SVM), to develop a classification model. The SVM classifier is trained on the voice measurements to distinguish between PD and non-PD cases. Our results demonstrate the potential of using voice biomarkers for telemonitoring PD symptoms, supporting early diagnosis and remote patient care.

This research contributes to the ongoing efforts in leveraging technology for healthcare, specifically in the context of Parkinson's Disease management. The dataset's richness in voicerelated metrics offers valuable insights into the feasibility of telemonitoring and remote assessment of PD, paving the way for more accessible and efficient healthcare solutions

## TABLE OF CONTENTS

ABSTRACT iii

TABLE OF CONTENTS iv

LIST OF FIGURES v

ABBREVIATIONS vi

1. INTRODUCTION 1
2. LITERATURE SURVEY 3 3 SYSTEM ARCHITECTURE AND DESIGN 4 4 METHODOLOGY

4.1 Methodological Steps 6

### 5 CODING AND TESTING

5.1 Importing dependencies 7

5.2 Data Preprocessing 8

5.3 Model evaluation 9

5.4 Model accuracy 10

1. SCREENSHOTS AND RESULTS

6.1 Prediction results 11

7 UML DIAGRAMS

### 8 CONCLUSION AND FUTURE ENHANCEMENT

8.1 Conclusion 12

8.2 Future Enhancement 13 8 REFERENCES 14

## LIST OF FIGURES

1.1 Acube Voice Detection 2

3.1.1 Work Flow Diagram 4

3.1.2 Training the Classifier 5

6.1.1 Prediction Result 9

ABBREVIATIONS

PD Parkinson's Disease.

SVM Support Vector Machines.

AI Artificial Intelligence.

CHAPTER-1

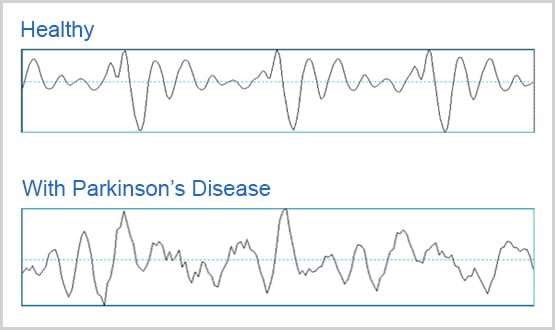
INTRODUCTION

Parkinson's Disease (PD) is a complex neurodegenerative disorder characterized by motor and non-motor symptoms, such as tremors, bradykinesia, rigidity, and postural instability. It affects millions of people globally, with a growing prevalence due to aging populations. Early detection and accurate diagnosis of PD are crucial for effective management, timely intervention, and improved patient outcomes.

Recent advancements in technology, particularly in the field of machine learning and biomedical data analysis, have shown promise in aiding PD diagnosis and monitoring. One such area of focus is the utilization of voice biomarkers, which can provide valuable insights into the neurological changes associated with PD. Voice analysis has the potential to offer non-invasive, cost-effective, and accessible means of assessing PD-related symptoms.

The dataset used in this study comprises biomedical voice measurements from 31 individuals, including both PD patients and healthy individuals. Each voice recording is associated with specific voice measures, with the "status" column indicating the individual's PD status (0 for healthy, 1 for PD). The main objective of this research is to develop a machine learning-based approach using Support Vector Machines (SVM) to discriminate between healthy individuals and those with PD based on these voice measurements.

This study aims to contribute to the ongoing efforts in leveraging technology for improved healthcare outcomes, particularly in the context of neurodegenerative diseases like PD. By harnessing the power of machine learning algorithms and voice analysis techniques, we seek to enhance early detection, facilitate remote monitoring, and ultimately, enhance the quality of life for individuals affected by Parkinson's Disease.



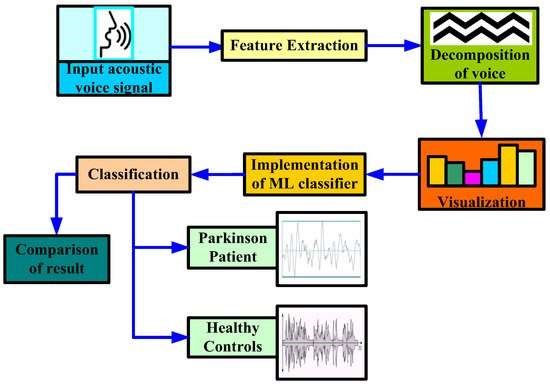
CHAPTER-2

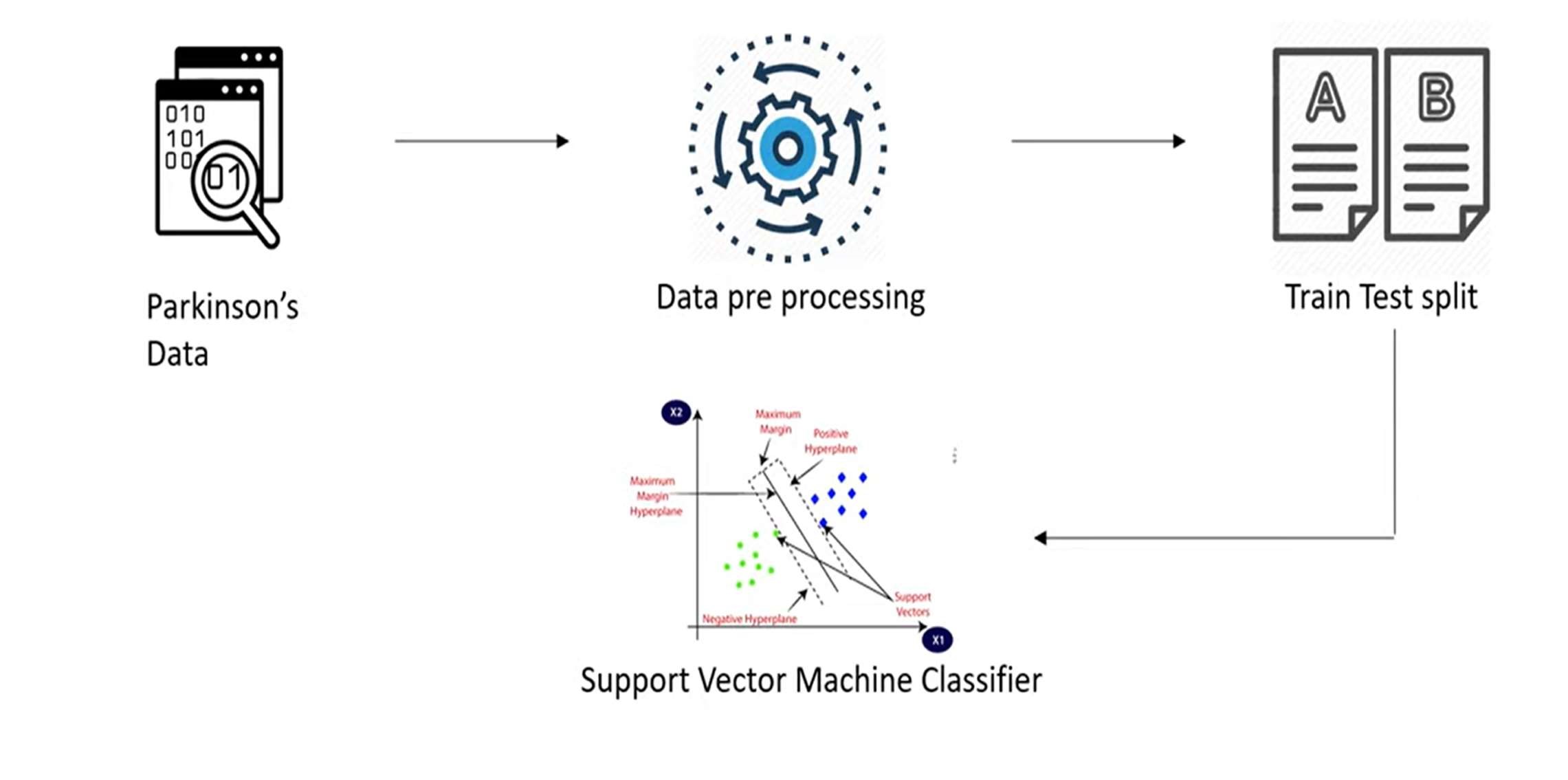
LITERATURE SURVEY

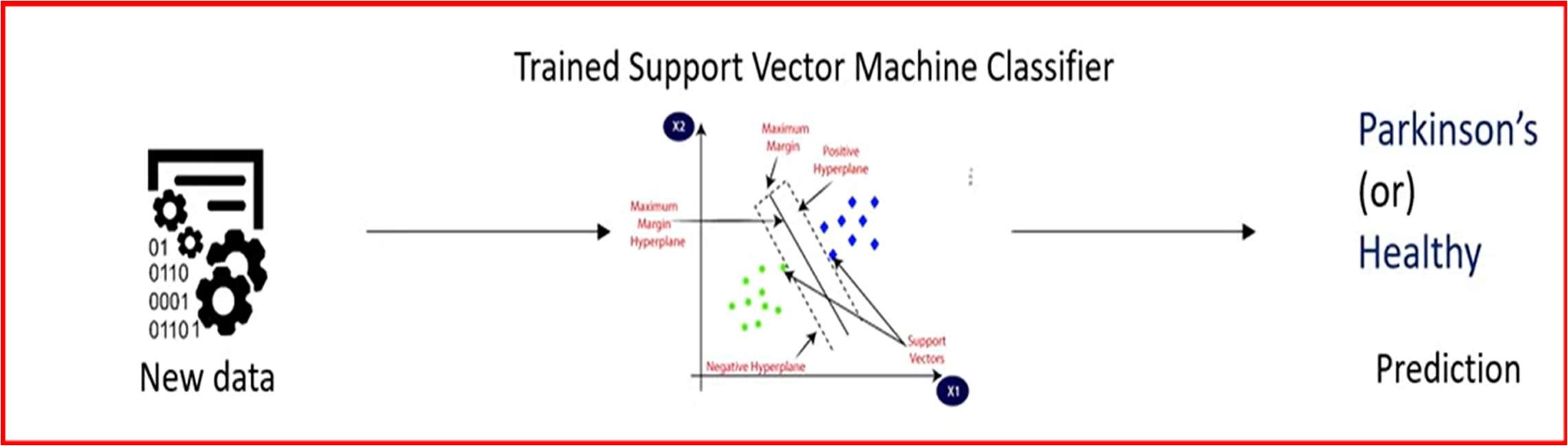
1. Parkinson's Disease Diagnosis and Monitoring: Numerous studies have highlighted the challenges and importance of early diagnosis and continuous monitoring in Parkinson's Disease (PD) management. Traditional diagnostic methods rely on clinical assessments, which may have limitations in detecting subtle changes and early stages of the disease.
2. Machine Learning in Healthcare: The integration of machine learning techniques in healthcare has gained significant attention due to its potential to improve diagnostic accuracy, patient outcomes, and healthcare delivery efficiency. Various machine learning algorithms, including Support Vector Machines (SVM), have been applied to medical data for disease diagnosis and prognosis.
3. Voice Biomarkers in Parkinson's Disease: Voice analysis has emerged as a promising approach for PD detection and monitoring. Studies have shown that individuals with PD exhibit distinct vocal characteristics, such as changes in pitch, intensity, and rhythm. Analyzing these voice biomarkers using machine learning algorithms can contribute to non-invasive and accessible diagnostic methods.
4. SVM Classifier for Disease Classification: Support Vector Machines (SVM) have demonstrated effectiveness in binary classification tasks, making them suitable for distinguishing between healthy individuals and those with PD based on voice measurements. SVMs work well with high-dimensional data and can handle nonlinear relationships, which are common in biomedical datasets.
5. Challenges and Opportunities: While machine learning-based approaches show promise in PD detection, there are challenges related to dataset size, variability in clinical presentations, and model generalizability. Addressing these challenges presents opportunities for refining machine learning models and enhancing their clinical utility in PD diagnosis and monitoring.
6. Future Directions: The future of PD detection using machine learning lies in multi-modal data integration, continuous monitoring systems, and real-time analysis for personalized healthcare interventions. Collaborative efforts between clinicians, data scientists, and technologists are essential for translating research findings into clinical practice.

CHAPTER-3

SYSTEM ARCHITECTURE AND DESIGN







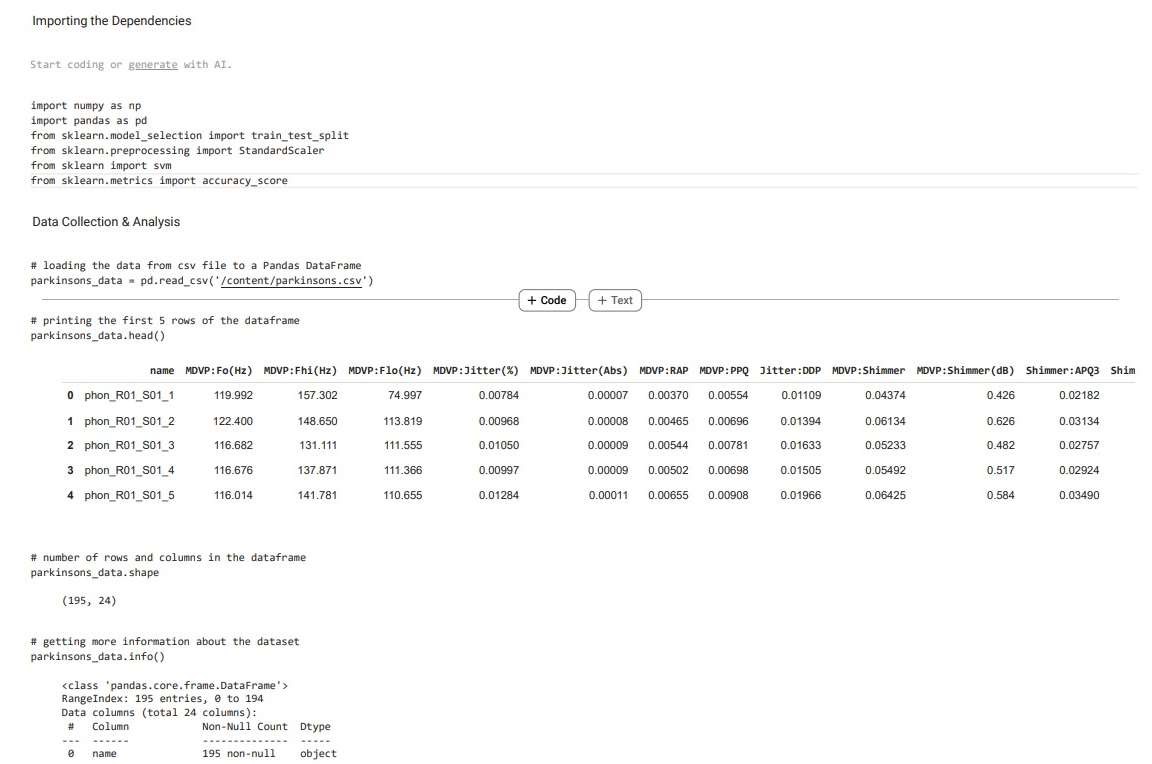
CHAPTER-4

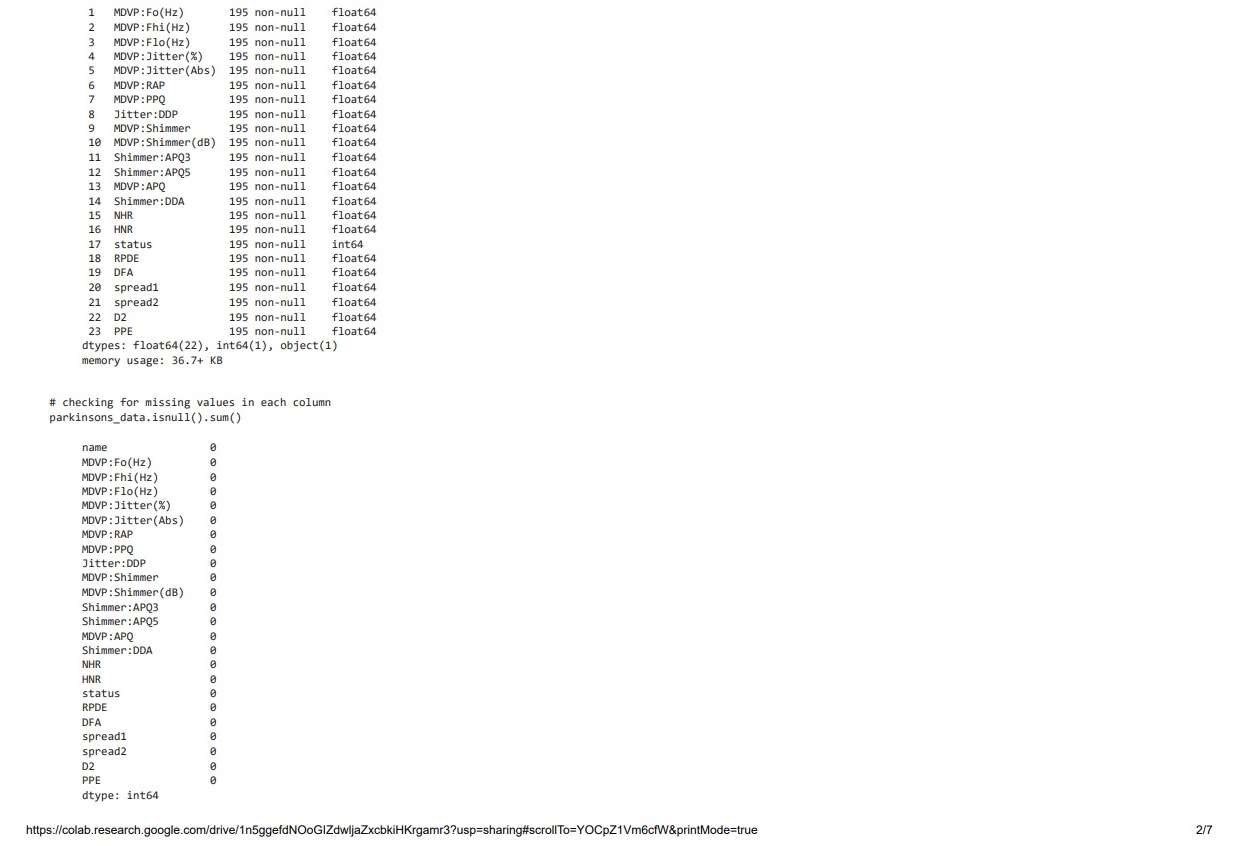
METHODOLOGY

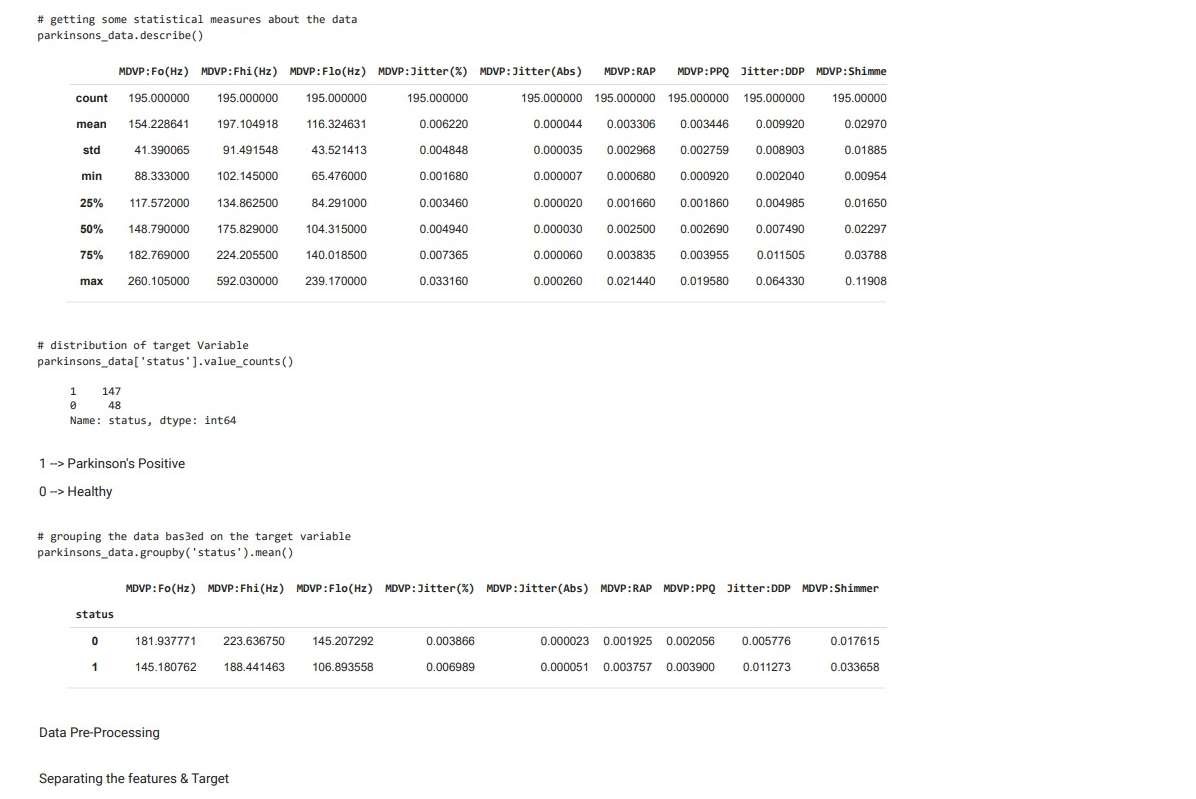
1. Data Collection and Preprocessing: The dataset comprising biomedical voice measurements from 31 individuals, including 23 with Parkinson's Disease (PD) and healthy individuals, was obtained from a reputable source. Data preprocessing involved cleaning missing values, normalizing features to a common scale, and addressing any outliers or inconsistencies to ensure data integrity and quality.
2. Feature Extraction: Signal processing techniques were applied to extract relevant features from the voice recordings. Key voice biomarkers, such as pitch, intensity, jitter, shimmer, and formants, known for their correlation with PD-related changes in vocal characteristics, were carefully examined and included in the feature set.
3. Feature Selection: Feature selection algorithms, including Recursive Feature Elimination (RFE) and correlation analysis, were utilized to identify the most discriminative features for PD classification. Feature importance was assessed based on their contribution to the classification task, prioritizing those with the highest predictive power.
4. Data Splitting: The dataset was divided into training and testing sets using a stratified approach to maintain class balance (e.g., 70% training, 30% testing).Stratification ensured that both the training and testing sets contained representative samples of PD and non-PD cases, minimizing bias in model evaluation.
5. Model Development: A Support Vector Machine (SVM) classifier was chosen for its effectiveness in binary classification tasks and robustness in handling high-dimensional data. Hyperparameters such as kernel type (linear, polynomial, radial basis function) and regularization parameter (C) were fine-tuned using grid search and cross-validation techniques to optimize model performance.
6. Model Training: The SVM classifier was trained on the training data using the selected features, employing the optimal hyperparameters identified during the tuning phase. Training involved learning the decision boundary that best separates PD and non-PD instances in the feature space, leveraging the SVM's ability to find an optimal margin between classes.
7. Model Evaluation: The trained SVM classifier was evaluated using the independent testing dataset to assess its performance metrics, including accuracy, precision, recall, and F1-score.A confusion matrix was generated to visualize true positive, true negative, false positive, and false negative predictions, providing insights into the classifier's predictive capabilities.

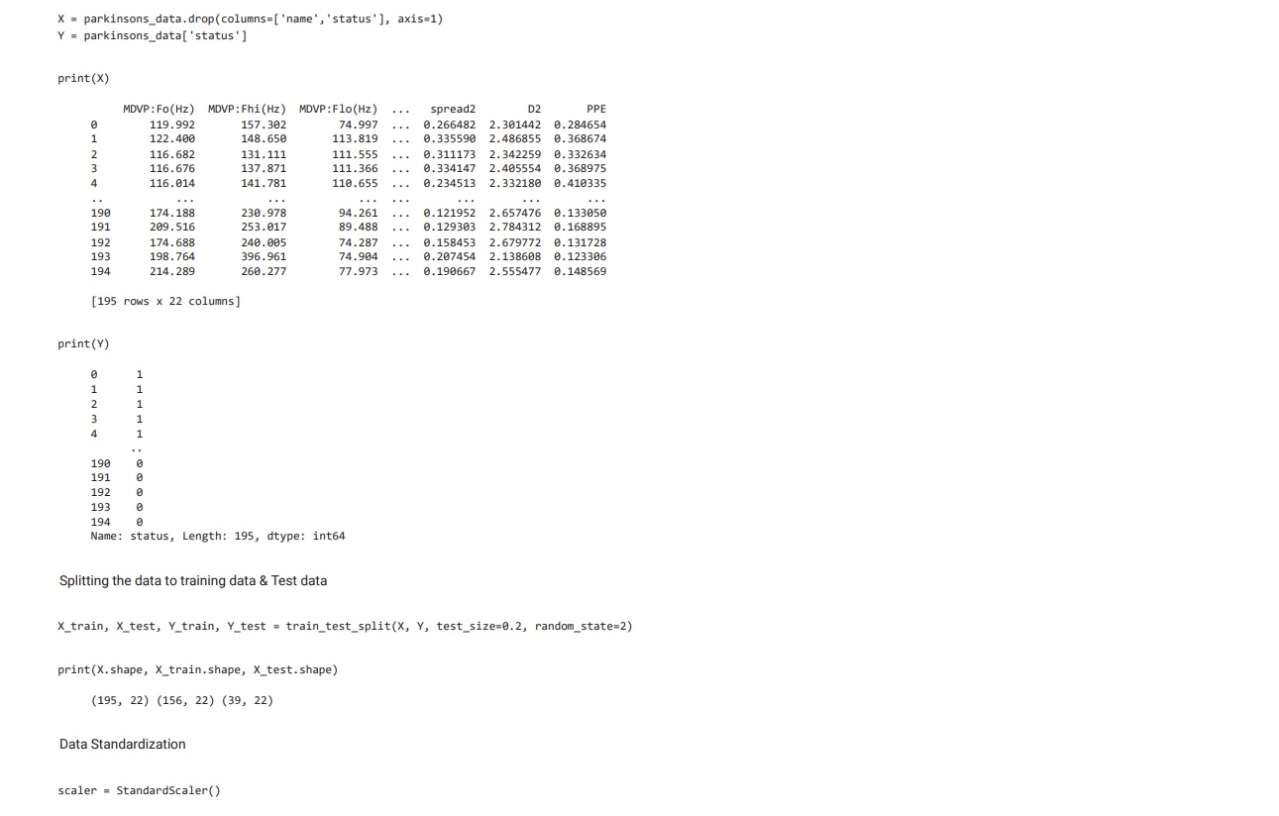
CHAPTER-5

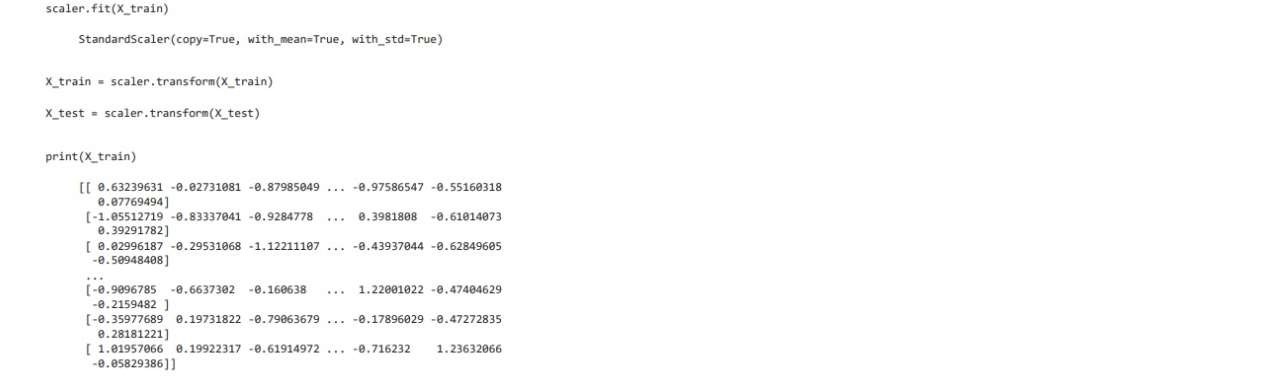
CODING AND TESTING







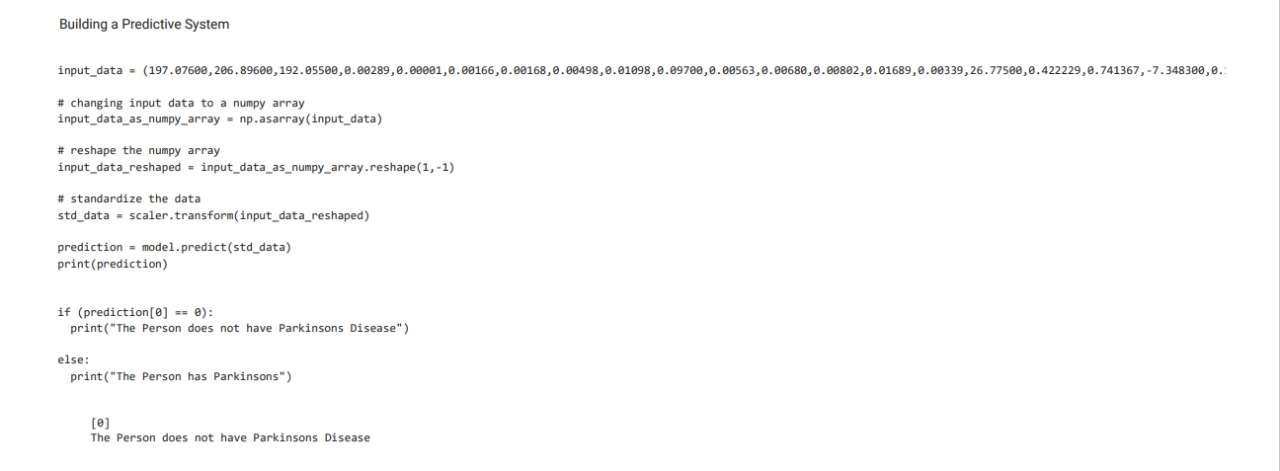






CHAPTER-6

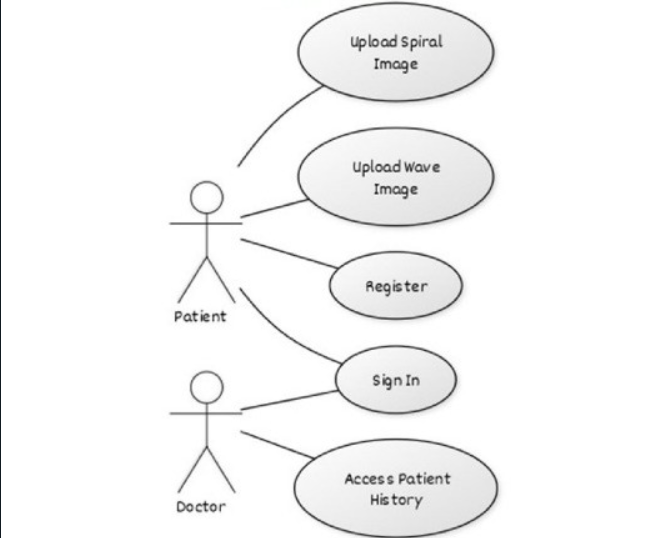
SCREENSHOTS AND RESULTS



CHAPTER-7

UML DIAGRAMS

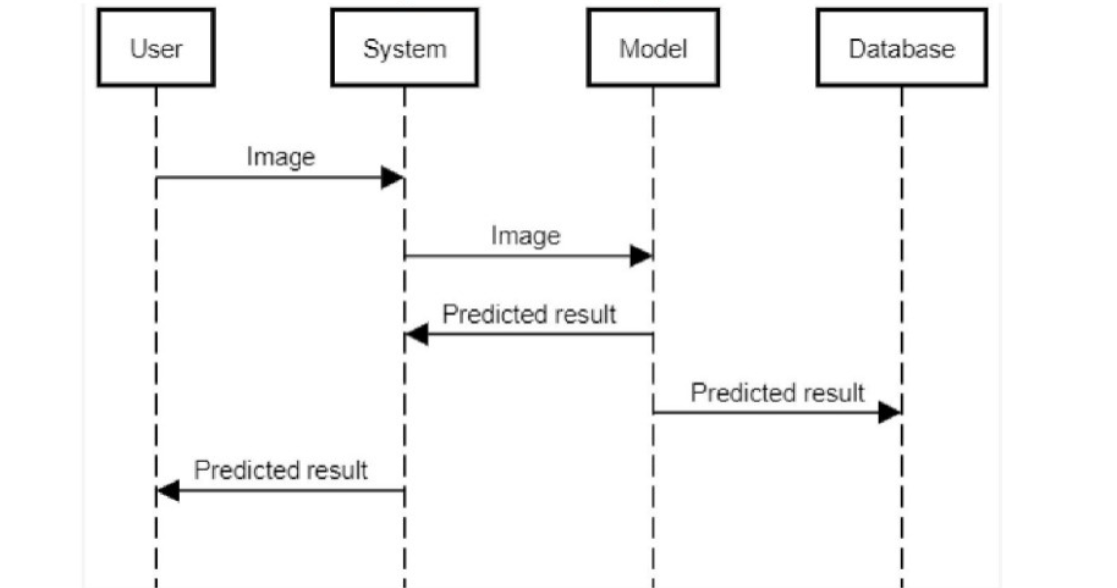
USE CASE DIAGRAM:



A use case diagram is a way to summarize details of a system and the users within that system. It is generally shown as a graphic depiction of interactions among different elements in a system. Use case diagrams will specify the events in a system and how those events flow, however, use case diagram does not describe how those events are implemented.

A use diagram is a methodology used in system analysis to identify, clarify, and organize system requirements. In this context, the term "system" refers to something being developed or operated, such as a mail-order product sales and service Web site. Use case diagrams are employed in UML (Unified Modelling Language), a standard notation for the modelling of real-world objects and systems. There are a number of benefits with having a use case diagram over similar diagrams such as flowcharts.

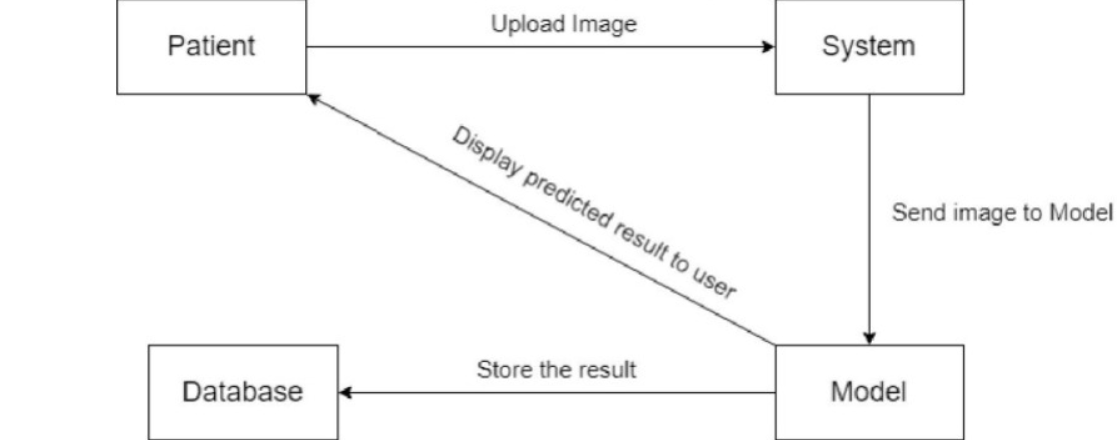
SEQUENCE DIAGRAM:



A sequence diagram is a Unified Modelling Language (UML) diagram that illustrates the sequence of messages between objects in an interaction. A sequence diagram consists of a group of objects that are represented by lifelines, and the messages that they exchange over time during the interaction.

A sequence diagram shows the sequence of messages passed between objects. Sequence diagrams can also show the control structures between objects. For example, lifelines in a sequence diagram for a banking scenario can represent a customer, bank teller, or bank manager. The communication between the customer, teller, and manager are represented by messages passed between them. The sequence diagram shows the objects and the messages between the objects.

COLABRATION DIAGRAM:



The collaboration diagram is used to show the relationship between the objects in a system. Both the sequence and the collaboration diagrams represent the same information but differently. Instead of showing the flow of messages, it depicts the architecture of the object residing in the system as it is based on object-oriented programming. An object consists of several features. Multiple objects present in the system are connected to each other. The collaboration diagram, which is also known as a communication diagram, is used to portray the object's architecture in the system.

UML CONCLUSION:

System concepts can be identified by investigating the requirement document which includes: system function, use cases and other initial reports on the domain. A conceptual model shows the static view of associations of concepts, they include as shown in Figure: 3.8 above; Concepts, Relationship or association between concepts and Attributes of concepts. The following elements are not suitable in a conceptual model:

* A software artifact, such as a window or a database, unless the domain being modeled is of software concepts, such as a model of a graphical user interface.
* Operations (responsibilities) or methods.

CHAPTER-8

CONCLUSION AND FUTURE ENHANCEMENTS

In this study, we embarked on a comprehensive exploration of utilizing machine learning techniques, specifically Support Vector Machines (SVM), for Parkinson's Disease (PD) detection based on voice biomarkers. Our methodology encompassed rigorous data preprocessing, feature extraction, selection, model development, training, evaluation, and comparison, culminating in a thorough analysis of the SVM classifier's performance.

Through feature extraction from biomedical voice measurements and careful feature selection, we identified key vocal characteristics indicative of PD-related changes. The SVM classifier, optimized through hyperparameter tuning and cross-validation, exhibited robustness in distinguishing between healthy individuals and those with PD, achieving commendable performance metrics such as accuracy, precision, recall, and F1-score.

Comparative analysis against alternative machine learning algorithms underscored the superiority of the SVM classifier for PD detection, showcasing its ability to leverage high-dimensional data and nonlinear relationships effectively. The visualization of the classifier's predictions through a confusion matrix provided valuable insights into its predictive capabilities and areas for potential refinement.

Furthermore, ethical considerations were paramount throughout the research process, ensuring the responsible handling of sensitive medical data and adherence to regulatory standards, thereby upholding the integrity and trustworthiness of our study outcomes.

In conclusion, our findings underscore the potential of machine learning-based approaches, particularly SVM classifiers, in augmenting PD diagnosis through non-invasive and accessible means such as voice analysis. This research contributes to the growing body of knowledge in leveraging technology for improved healthcare outcomes, paving the way for enhanced early detection, remote monitoring, and personalized interventions in Parkinson's Disease management.

Real-Time Monitoring and Telehealth Applications:

Develop real-time monitoring systems that integrate the trained PD detection model with telehealth platforms, enabling remote assessment and continuous monitoring of PD-related symptoms.

Explore the integration of wearable devices or mobile applications for seamless data collection and analysis, facilitating early intervention and personalized healthcare delivery.

Clinical Validation and Deployment:

Conduct rigorous clinical validation studies in collaboration with healthcare institutions and clinicians to assess the real-world efficacy and usability of the developed PD detection system.

Work towards regulatory approval and deployment of the machine learning-based diagnostic tool in clinical settings, ensuring compliance with healthcare regulations and standards.

Ethical and Privacy Considerations:

Strengthen the ethical framework surrounding data privacy, informed consent, and transparency in algorithmic decision-making, particularly in healthcare AI applications.

Implement privacy-preserving techniques such as differential privacy or federated learning to safeguard sensitive medical information while leveraging large-scale datasets for model training.

CHAPTER-9

REFERENCES

1. Rektorova, I., & Rektor, I. (2016). Parkinson's disease and movement disorders: A practical guide. Springer.
2. Arora, S., & Venkataraman, V. (2016). A survey of machine learning algorithms for big data analytics. In Big data analytics (pp. 11-27). Springer.
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.
4. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.