

**EARLY DETECTION OF AlZHEIMER’S DISEASE**

**A MINI PROJECT REPORT**

***Submitted by***

**AROKIA NADHIA V (1920110004)**

**JOE HAYACINTH MACBHEL D A (1920110018)**

**SHAANAVI DEVI S (1920110043)**

***in partial fulfillment for the award of the degree***

***of***

**BACHELOR OF TECHNOLOGY**

**IN**

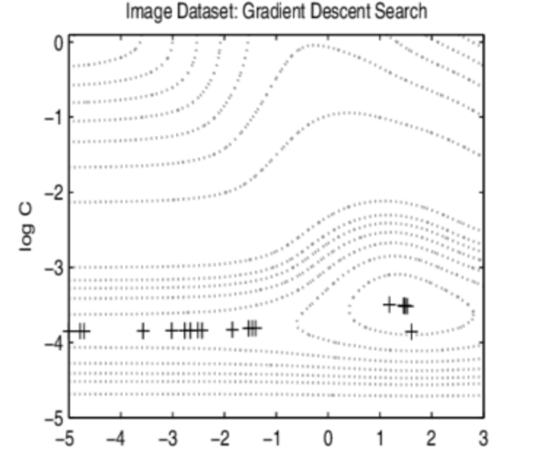
ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

**SONA COLLEGE OF TECHNOLOGY, SALEM-5**

**(Autonomous)**

**ANNA UNIVERSITY: CHENNAI 600 025**

November 2023



**ANNA UNIVERSITY: CHENNAI 600 025**

**BONAFIDE CERTIFICATE**

Certified that this project report “**EARLY DETECTION OF ALZHEIMER’S DISEASE”** is the bonafide work of **“AROKIA NADHIA V 1920110004”, “JOE HAYACINTH MACBHEL D A 1920110018”, “SHAANAVI DEVI S 1920110043”** who carried out the project work under my supervision.

**SIGNATURE**

**SIGNATURE**

Dr. J. Akilandeswari

C.Santhosh Kumar

**HEAD OF THE DEPARTMENT**

**SUPERVISOR**

Professor

Assistant Professor

Department of Information Technology

Department of Information Technology

Sona College of Technology

Sona College of Technology

Salem- 636 005.

Salem- 636 005.

Submitted for Project viva voce examination held from 16.11.2023 to 18.11.2023

|  |  |
| --- | --- |
| **INTERNAL EXAMINER** | **EXTERNAL EXAMINER** |

i

**ACKNOWLEDGEMENT**

First and foremost, we thank to **power of almighty** for showing us inner peace and for all blessings .Special gratitude to our parents, for showing their support and love always.

We express our sincere thanks to chairman **Sri.C.VALLIAPPA** and principal **Dr.S.R.R.SENTHILKUMAR** for providing adequate facilities to complete the project.

We express our sincere thanks to Head of the Department **Dr. J. AKILANDESWARI** for her unwavering support in driving our team's motivation and dedication throughout the project.

We express our heartfelt thanks to our project supervisor **C.SANTHOSH KUMAR** for his valuable guidance and fruitful discussions throughoutthe course of the project work.

We feel proud in sharing this success with all our department Faculty, Staff members and friends who helped directly and indirectly to completing this project successfully.

V.AROKIA NADHIA D.A. JOE HAYACINTH MACBHEL

S.SHAANAVI DEVI

ii

**ABSTRACT**

Alzheimer's disease is a major cause of cognitive impairment in adults at age more than 65, which leads to high treatment costs and contributes to elderly mortality. Early detection using machine learning is vital to reduce risks. This research aims to create a computer aided Alzheimer's detection system. Neurodegenerative changes in brain affect memory and function, posing significant social and economic challenges. Early analysis is more effective. We have used machine learning techniques such as support vector machine with Oasis data set to predict Alzheimer's evaluated by precision recall accuracy and F1 score.This research involves a comprehensive data set making up Alzheimer Patients mild cognitive impairment individuals and healthy controls this data set. Here classifiers were trained and tested from the database that we have showed and even several classifiers were compared and combined for a decision through voting. To achieve this, we have implemented machine learning and Python using SVM models. In conclusion predictions of Alzheimer’s disease are based on the measured parameters like accuracy, precision, recall for support vector machine (SVM) model. The proposed classification scheme can be used by doctors to make diagnosis of Alzheimer disease based on the accuracy from the output.

iii

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CHAPTER NO.** | **TITLE** |  | | **PAGE NO.** | |
|  | **ABSTRACT** | | | **iii** | |
|  | **LIST OF FIGURES** | | | **v** | |
|  | **LIST OF TABLES** | | | **vi** | |
|  | **LIST OF ABBREVIATIONS** | | | **vii** | |
| **1.** | **INTRODUCTION** | | | **1** | |
|  | 1.1. | About the Project | | 1 | |
|  | 1.2. | Role in the Project | | 2 | |
|  | 1.3. | Scope of the Project | | 4 | |
|  | 1.4. | Literature Survey | | 5 | |
|  | 1.5. | Existing System | | 7 | |
| **2.** | **SYSTEM DESIGN** | | | **10** | |
|  | 2.1. | Dataset Preparation | | 10 | |
|  | 2.2. | Algorithm used | | 12 | |
|  | 2.3. | Methodology | | 15 | |
|  | 2.4. | Modules | | 16 | |
|  | 2.5. | Screenshots and Results | | 17 | |
|  | 2.6. | Comparison with other Algorithm | | 19 | |
| **3.** | **CONCLUSION AND FUTURE WORKS** | | | **23** | |
|  | 2.5. | | Conclusions | | 23 | |
|  | 2.6. | | Future Works | | 24 | |
|  | **REFERENCES** | | | **25** | |
|  | **ONLINE REFERENCES** | | | **26** | |
|  |  | | |  | |

iv

**LIST OF FIGURES**

|  |  |
| --- | --- |
| **NAME OF THE FIGURE** | **Page No** |
| Fig 1.1 Four kinds of Alzheimer’s disease | 2 |
| Fig 1.2 Cleaning process of large datasets | 3 |
| Fig 1.3 Existing System | 8 |
| Fig 2.1 Dataset preparation | 9 |
| Fig 2.2 Detection of disease | 10 |
| Fig 2.3 Shrunk image of brain | 10 |
| Fig 2.4 Pychological parameters | 11 |
| Fig 2.5 SVM Model | 12 |
| Fig 2.6 Featues selection process | 13 |
| Fig 2.7 Methodology | 14 |
| Fig 2.8 Four kinds of Dementia in Early detection | 15 |
| Fig 2.9 Dementia rate for gender | 16 |
| Fig 2.10 Density rate of disease | 17 |
| Fig 2.11 Comparison between dementia  and non dementia | 18 |
| Fig 2.12 EDUC level | 18 |
| Fig 2.13 Graph plot | 19 |
| Fig 2.14 Logistc regression and SVM | 20 |
| Fig 2.15 Comparison between SVM and Random forest | 21 |
| Fig 2.16 Comparison between Algorithms | 22 |

**v**

**LIST OF TABLES**

|  |  |
| --- | --- |
|  |  |
| Table 1.1 Comparison between Algorithms | 24 |

**vi**

**LIST OF ABBREVIATIONS**

**SVM-** Support vector machine

**API-**Application Program Interface

**ML-** Machine learning

**MRI-**Magnetic Resonance Imaging

**MCI-**Mild Cognitive Impairment

**PET-** Positron Emission Tomography

**CSF-** Cerebro Spinal Fluid

**AD**-Alzheimer’s Disease

**FN**-False Negative

**FP**-False Positive

**TN**-True Negative

**TP**-True Positive

**CV**-Cross Validation

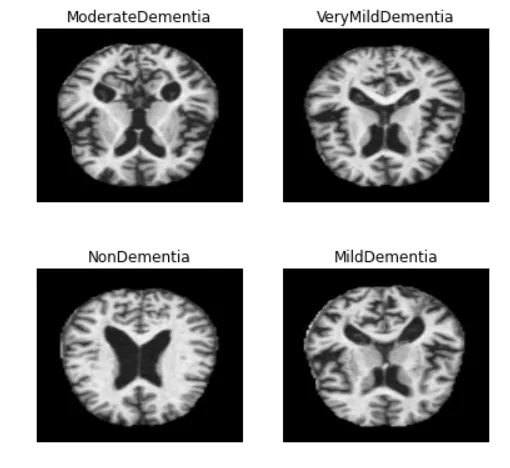
**vi**

**CHAPTER 1**

**INTRODUCTION**

**1.1 ABOUT THE PROJECT**

Alzheimer's disease is a progressive neurological condition causing short term memory loss affecting around 5.1 million people in US with no cure. So we are analysing the early stage of Alzheimer’s diseases using SVM Model and also we have used MRI images as the primary dataset for this research project. This MRI scan is being thrown into a web application which will analyse whether the particular individual is suffering from this disease or not based on the four kinds of images being discussed below. Automated system more exact than humans are used in medical decision support fill stop various data sources, including MRI scans and biomarkers helps to predict AD.We have used Support vector machine that to predict Alzheimer's evaluated by precision recall accuracy and F1 score. Early-stage AD allows independence, but memory issues persist. As AD progresses, communication and daily activities becomes challenging. This affects mostly the elders this disease is a generic form of dementia that accounts for 60-80% of dementia cases. Machine learning models offer hope for early detection by analysing data from MRI scans and other sources to stop this research aims to create a computer aided casinos detection system using machine learning.These changes will have profound social financial and economic implications. Therefore, in this research we have differentiated four types like, Demented, Non-Demented, Moderate demented, Very Mild demented images that helps the doctors to predict preliminary stages of Alzheimer's disease.The four kinds of images are shown in the Figure.1.1 This will in turn analysis the accuracy, precision, and recall along with the F-score. This research involves a comprehensive data set making up Alzheimer Patients mild cognitive impairment individuals and healthy controls this data set. Here classifiers were trained and tested from the database that we have showed and even several classifiers were compared and combined for a decision through voting.



**Figure.1.1**.Four kinds of Alzheimer’s disease.

**1.2 ROLE IN THE PROJECT**

We divided the role into four in order to assume good results and they are as follows:

**1.2.1 MACHINE LEARNING ENGINEER**

Being an Machine learning engineer, we are responsible for developing and fine tuning the SVM model.we will be handling cross-validation and hyper parameter tuning to ensure model accuracy.As a ML Engineer, I will be preprocessing and analysing medical data like MRI images with the help of SVM models. As a technical consultant, our role is to translate those functional documents into technical documents and write code to implement business needs, developing objects.

**1.2.2 DATA ANALYST**

Being an Data analyst, we will be cleaning the large datasets including clinical data (MRI),neuroimaging and genetic data and also we will be collaborating with the data scientist to ensure the data quality and integrity.Figure.1.2 ensures the cleaning process of large MRI datasets. The final analysis is made and proceded with the further process . Also we assists in data collection and initial data exploration and may perform basic data preprocessing as discussed above.

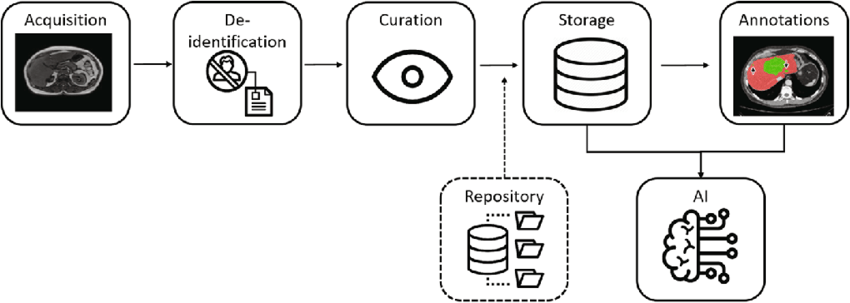
**1.2.3 PROJECT MANAGER**

Being a Project manager, we will act as a liason between stakeholders and technical experts, including healthcare professionals. we will bulid and collect the data and storage that we have obtained from the process.

**1.2.4 QUALITY ASSURANCE SPECIALIST**

Being a Quality Assurance Specialist,we willensure the accuracy and the reliability of the data and models.And also,we will conduct thorough testing and validation to identify and correct any issues.

And also we will collaborate in the clinical assessment and diagnosis of Alzheimer’s disease patients.We will be consulting good quality assurance specialist for the process .Finally, we will ensure products meet quality standards and in medicinal fields by monitoring process and improvement to enhance operations and ensuring reliability of systems and minimize downtime.



**Figure.1.2.**Cleaning process of large dataset**.**

**1.3.SCOPE OF THE PROJECT**

The project's scope, centered on Priliminary assessment of Alzheimer's disease through SVM model, aims to create a predictive system for the early detection and continuous monitoring of Alzheimer's disease. This undertaking involves gathering and analyzing a wide range of patient data, encompassing medical records, cognitive evaluations, and medical images, to construct a robust SVM model. This model will be trained to identify subtle cues indicative of Alzheimer's disease in its early stages by recognizing patterns and correlations in the data. The overarching objective is to furnish healthcare professionals with a valuable tool for prompt intervention and tailored patient care, potentially enhancing the quality of life for individuals at risk of or already afflicted by Alzheimer's disease. Furthermore, this project can make a valuable contribution to Alzheimer's research by pinpointing new biomarkers and risk factors associated with the condition.

In addition to the primary objectives outlined above, this project can also have several supplementary dimensions. Firstly, it could significantly reduce the burden on healthcare systems by streamlining the diagnostic process, making it more efficient and cost-effective. Early detection and intervention may lead to a reduction in the long-term healthcare costs associated with Alzheimer's disease. Furthermore, the insights gained from the project may lead to a deeper understanding of the disease's progression and its interaction with various risk factors, thus paving the way for more targeted therapeutic approaches and potential preventive measures. Collaborations with academic institutions and healthcare organizations can foster the sharing of data, which is crucial for developing robust models. Moreover, this initiative might encourage further research into the ethical and privacy considerations surrounding the use of sensitive health data in machine learning applications, contributing to the broader discourse on responsible AI in healthcare. Ultimately, the project's impact extends beyond diagnostics, potentially transforming the landscape of Alzheimer's disease research, care, and policy.

**1.4.LITERATURE SURVEY**

**Sivakani GA, Ansari R. (2020) “Machine learning framework for implementing Alzheimer's disease*.*”**

Numerous studies have explored Alzheimer's disease (AD) detection using machine learning. Researchers have employed diverse datasets, including OASIS, to develop predictive models. Various machine learning algorithms, such as support vector machines, decision trees, and neural networks, have been applied for early AD diagnosis. Feature selection techniques, like Recursive Feature Elimination (RFE), have been crucial in improving model performance. These studies emphasize the importance of early detection to mitigate brain cell degeneration, focusing on features related to cognitive, genetic, and neuroimaging data. Overall, the literature underscores the potential of machine learning in aiding early AD diagnosis and enhancing our understanding of the disease.

**Kavitha C, Mani V, Srividhya SR, Khalaf OI, Tavera Romero CA.(2022) “Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. Front Public Health.”**

Alzheimer's disease (AD) is a pressing concern as it leads to dementia in older adults, impacting individuals, families, and healthcare systems worldwide. The rising incidence rates of AD and other metabolic diseases like diabetes necessitate innovative solutions. Machine learning techniques, including Decision Trees, Random Forests, Support Vector Machines, Gradient Boosting, and Voting Classifiers, have been applied to predict AD using Open Access Series of Imaging Studies (OASIS) data. Early diagnosis is crucial, as treatment is more effective and less damaging in the initial stages. These models exhibit promising results, with a validation average accuracy of 83%, outperforming existing approaches.

**Khalaf OI, Abdulsahib GM, Sabbar BM. (2020) “Optimization of wireless sensor network coverage using the Bee Algorithm. J Inf Sci Eng.”**A method for optimizing the wireless sensor network coverage was introduced using the bee algorithm. We simulated the proposed method in MATLAB software and compared the obtained results with the genetic algorithm. The results showed that the bee algorithm gives more optimal coverage percentage compared to the genetic algorithm and uses less time to use the system resources and implement the algorithm.

**Sudharsan M, Thailambal G. (2021). “Alzheimer's disease prediction using machine learning techniques and principal component analysis (PCA), Materials Today: Proceedings.”**Alzheimer's disease is a debilitating neurodegenerative condition with no known cure, making early detection crucial for effective management. However, the scarcity of trained samples and extensive feature descriptions pose challenges for accurate diagnosis. This study introduces a novel method using Mild Cognitive Impairment (MCI) and Structural Magnetic Resonance (sMR) imaging to differentiate between Alzheimer's patients and healthy controls. It employs Import Vector Machine (IVM), Regularized Extreme Learning Machine (RELM), and Support Vector Machine (SVM) with a score-based feature selection approach and dynamic data transformations. Results from Alzheimer's Disease Neuroimaging Initiative (ADNI) datasets demonstrate that RELM significantly improves AD classification accuracy, benefitting early diagnosis and treatment.

**Saratxaga CL, Moya I, Picón A, Acosta M, Moreno-Fernandez-de-Leceta A, Garrote E, et al(2021). “MRIDeep learning-based solution forAlzheimer's Disease Prediction. J.Pers. Med*.*”**Alzheimer's disease is a progressive, debilitating condition, and early diagnosis is crucial for timely intervention and treatment. While tests like the Mini-Mental State Examination are common for early identification, magnetic resonance imaging (MRI) brain analysis is essential .

This study presents a novel approach using deep learning and image processing techniques for MRI-based Alzheimer's diagnosis, surpassing previous research. The method achieves a balanced accuracy (BAC) of up to 0.93 for automated disease diagnosis and a BAC of 0.88 for disease staging. These results, based on the OASIS dataset, highlight the effectiveness of deep learning in creating a robust Alzheimer's diagnosis tool using MRI data.

**1.5 EXISTING SYSTEM**

The existing system of early analysis of Alzheimer's disease using SVM is still under development, but it has the potential to be a valuable tool for clinicians. The system works by training an SVM model on a dataset of neuroimaging data from Alzheimer's patients and healthy controls. The trained model can then be used to predict the probability that a new patient has Alzheimer's disease.

One of the most promising aspects of SVM-based systems for the early analysis of Alzheimer's disease is that they can be trained to identify patterns in complex data sets that are difficult to detect with the naked eye. For example, SVM models have been shown to be able to identify subtle changes in the volume of different brain regions that are associated with Alzheimer's disease. Another advantage of SVM-based systems is that they can be used to develop non-invasive and relatively inexpensive diagnostic tools. For example, SVM models have been trained to predict the probability of Alzheimer's disease based on MRI data. This means that SVM-based systems could be used to diagnose Alzheimer's disease in patients without the need for more invasive procedures, such as brain biopsies.

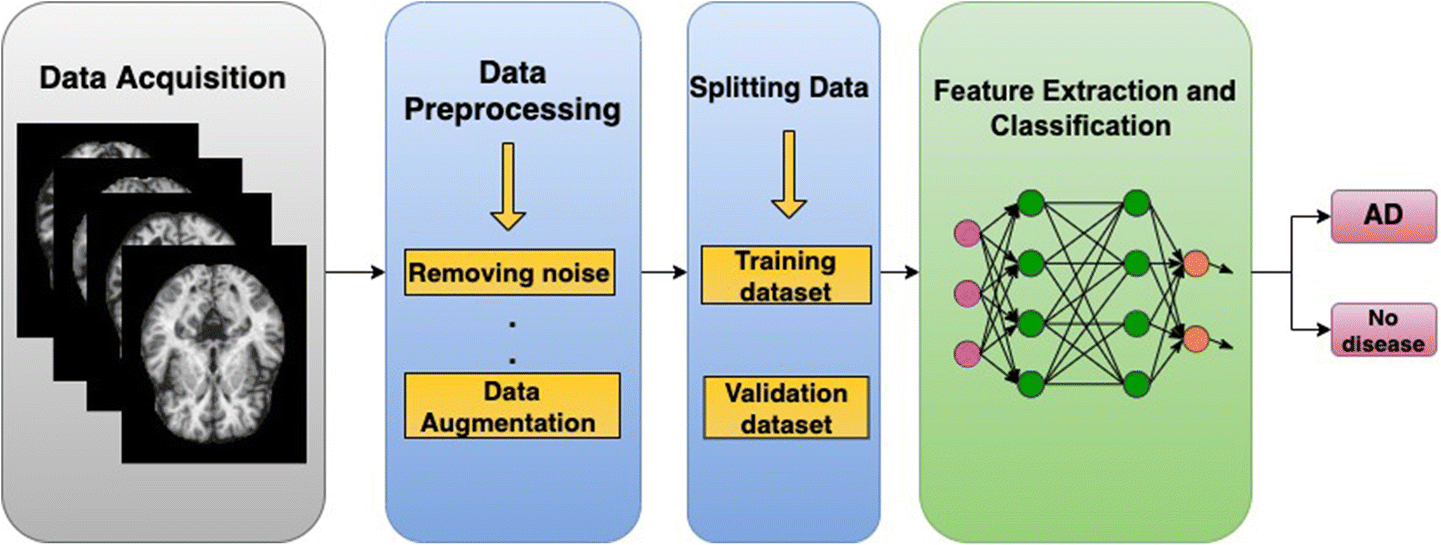
Despite the promise of SVM-based systems for the early analysis of Alzheimer's disease, there are still some challenges that need to be addressed before these systems can be widely adopted in clinical practice. One challenge is that SVM models can be computationally expensive to train, especially on large datasets. Another challenge is that SVM models can be sensitive to the choice of hyperparameters. It is important to carefully select the hyperparameters to optimize the performance of the model. Finally, it is important to ensure that SVM models are interpretable.

Overall, SVM-based systems for the early analysis of Alzheimer's disease have the potential to revolutionize the way that Alzheimer's disease is diagnosed and treated. By enabling the early detection of Alzheimer's disease, SVM-based systems can help clinicians to intervene earlier and improve the quality of life for patients with Alzheimer's disease.

Here are some examples of existing systems for the early analysis of Alzheimer's disease using SVM:

* The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal study that is collecting neuroimaging data from Alzheimer's patients and healthy controls. The ADNI data has been used to train SVM models that can predict the probability of Alzheimer's disease with high accuracy.
* The Alzheimer's Disease Prediction Algorithm (ADPA) is a commercial software package that uses SVM to predict the probability of Alzheimer's disease based on MRI data. The ADPA has been shown to be effective in predicting Alzheimer's disease in patients with mild cognitive impairment (MCI).

These are just a few examples of the existing research and development efforts in the area of SVM-based systems for the early analysis of Alzheimer's disease. As the research continues and the technology matures, we can expect to see SVM-based systems play an increasingly important role in the diagnosis and treatment of Alzheimer's disease.Figure 1.3 shows the Existing system of the process.



**Figure 1.3** Existing System

**CHAPTER 2**

**SYSTEM DESIGN**

**2.1 DATASET PREPARATION**

This dataset consists of MRI images, where the dataset should include images with Demented, Moderate, very mild, and non-demented images. The dataset should be sufficiently large to train and evaluate SVM models efficiently. We have been using longitudinal MRI data for our analysis, which includes records from 150 individuals aged between 60 and 96. This binary classification is essential for supervised learning with SVM. This dataset encompasses longitudinal MRI records for each subject, with every participant having been scanned at least once. Preprocessing the data should be done to handle the missing data. This helps ensure that the model's accuracy is consistent across different data partitions. We have also plotted graph variations according to the gender and age.

The below figure shows the dataset process of the early prediction of Alzheimer’s disease. Among the subjects, X of them were consistently categorized as 'Nondemented' during the entire study duration. Additionally, Y subjects were initially classified as 'Demented' at their first visit and supported this classification throughout the study. Along with this, A and Z of them were categorised as ‘very mild’ and ‘moderate’ during the study duration. A common split ratio of 70% for training,15% for validation,15% of testing. Randomization is crucial to ensure unbiased model evaluation.

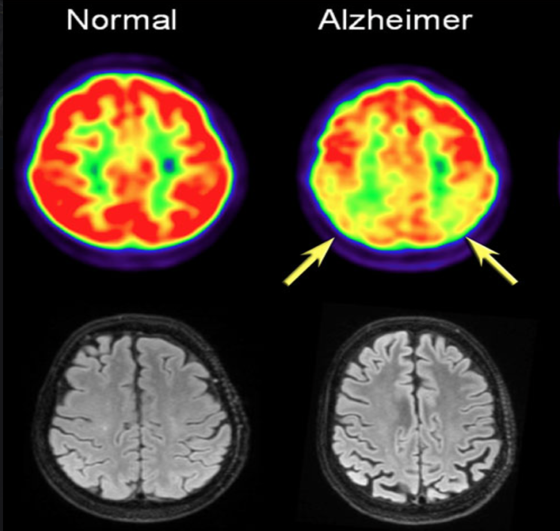
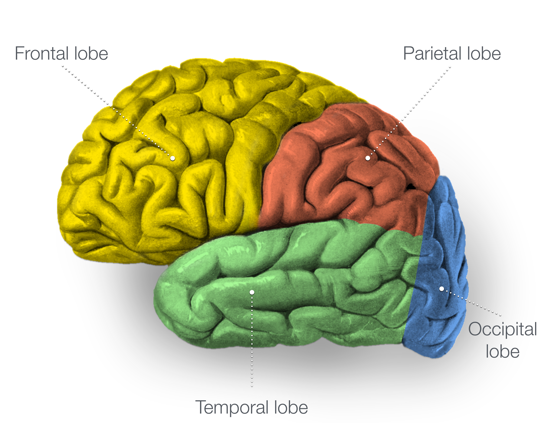
A diagram of a data processing process

Description automatically generated

**Figure 2.1** Dataset Preparation

**MRI**:

MRI is a non-invasive medical imaging technique which supplies detailed images of the brain’s structure and anatomy. It uses strong magnetic fields and radio waves to generate high-resolution. We have used MRI data for Alzheimer’s detection that is collected through scanning individuals, both those with Alzheimer’s detection is collected through scanning individuals, both those with Alzheimer’s disease and healthy controls, in medical imaging facilities. The dataset is split into three subsets, namely training, validation, and testing. MRI data is a crucial tool in Alzheimer's detection and research, supplying detailed insights into brain structure and function.

****

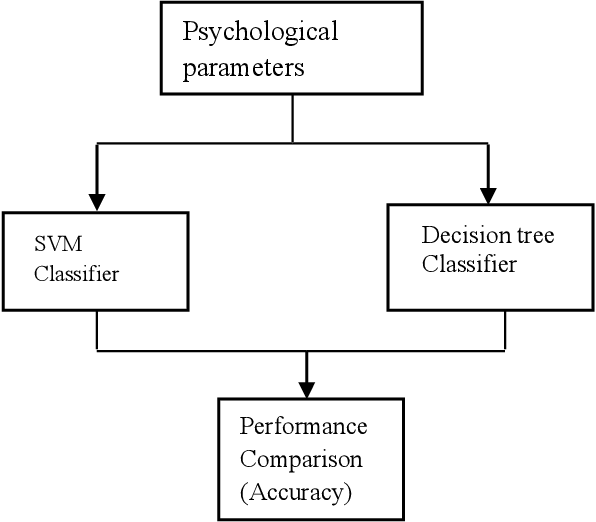
**Figure 2.2** Detection of Disease. **Figure 2.3** Shrunk image of brain

The parietal lobe (red shaded region) of the brain in the Figure 2.2 is the area which is used to find and detect Alzheimer’s disease. Figure 2.3 can show whether areas of the brain have atrophied (shrunk) which detects Alzheimer’s disease. Magnetic Resonance Imaging (MRI) images are an indispensable resource for Alzheimer's disease prediction and diagnosis due to their ability to capture a wide range of critical information about the brain. These images not only reveal structural changes but also provide insights into the brain's functional and metabolic aspects.Combining these various MRI modalities allows for a more comprehensive understanding of the disease's progression, aiding in the development of more accurate and personalized predictive models.

**2.2 ALGORITHM USED**

A crucial algorithmic approach is employed to detect and predict the onset of the disease. Researchers utilize MRI (Magnetic Resonance Imaging) images as a valuable dataset for this purpose. SVM models are particularly well-suited for this task due to their effectiveness in binary classification problems, making them ideal for distinguishing between Alzheimer's and non-Alzheimer's cases. To implement this algorithm, the MRI images are pre-processed to extract relevant features that capture structural and morphological information about the brain. These features are then used to train the SVM model.

This learns to classify patients into two categories: those with Alzheimer's and those without. The SVM model uses a hyperplane to separate the two classes, maximizing the margin between them, which enhances the model's generalization and predictive accuracy. This algorithmic approach proves to be a powerful tool in the preliminary analysis of Alzheimer's disease, offering the potential for timely interventions and improved patient care. Machine learning techniques like Support vector machines (SVM), and decision tree plays a pivotal role in this diagnostic processes. Figure 2.4 shows the Psychological parameters of this diagnostic process.These parameters encompass cognitive and behavioral aspects, including memory, attention, language skills, mood, and quality of life, among others. By integrating these psychological parameters with clinical and demographic data, machine learning models can be trained to recognize subtle changes that occur in the early stages of the disease. SVM and Decision Tree algorithms excel in processing and classifying this multidimensional information, making them valuable tools for predicting Alzheimer's disease onset. These models can discern patterns and relationships within the data, aiding in the early identification of individuals at risk. This approach not only enables timely interventions but also enhances our understanding of the disease's progression, ultimately contributing to more effective management and care for individuals affected by Alzheimer's disease.

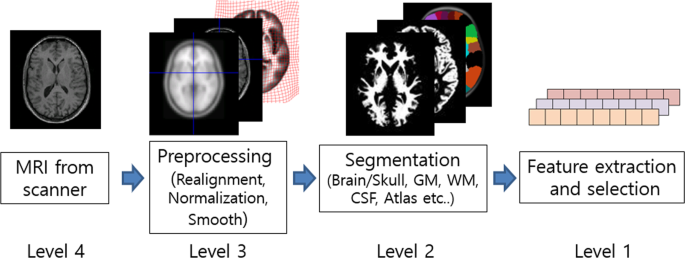


**Figure 2.4** Psychological Parameters

**2.2.1 SVM**

Support Vector Machine (SVM) plays a vital role in the early analysis of Alzheimer's disease using machine learning, particularly when applied to SVM models. Alzheimer's disease, a progressive neurodegenerative condition, is known for its subtle onset, making early diagnosis a significant challenge. This is where SVMs come into play. SVM is a supervised machine learning algorithm that is used for classification and regression tasks. In the context of Alzheimer's disease, SVM is primarily employed as a classification tool, aiming to distinguish between individuals who may develop Alzheimer's and those who may not based on various features and biomarkers.Figure 2.5 shows about the SVM model that we have used in this process. One of the primary advantages of SVMs is their capability to handle complex, non-linear data patterns effectively. In Alzheimer's disease analysis, SVM can learn and identify intricate relationships between a multitude of variables, including clinical data, neuroimaging results (such as MRI or PET scans), and genetic information. By using SVMs, researchers can combine these diverse data sources to develop a comprehensive diagnostic tool. The essence of SVM lies in its ability to find an optimal hyperplane that best separates different classes of data points. In the case of Alzheimer's disease, it can learn to distinguish between patients who will develop the disease and those who won't, or between different stages of cognitive decline. The SVM algorithm aims to maximize the margin between the data points of different classes, which essentially means finding the widest possible gap between them.

Moreover, the SVM model is versatile and can be continually improved through iterative processes. Regular updates can incorporate new data and emerging biomarkers, ensuring the model remains relevant and accurate for early Alzheimer's disease analysis. This adaptability is essential in a field where our understanding of the disease is continuously evolving, and new diagnostic tools are being developed.



**Fig 2.5** SVM Model

**2.3 METHODOLOGY**

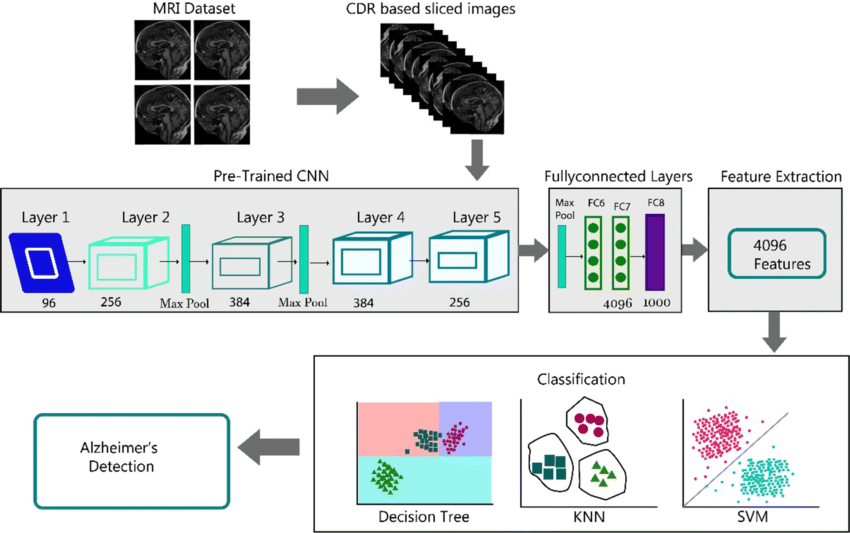
The method for early analysis of Alzheimer’s disease using SVM model typically involves several key steps and are shown in the Figure.2.7.

**2.3.1 DATA COLLECTION**

Gathers comprehensive datasets including genetic, medical, lifestyle, and cognitive information from a diverse population. By obtaining all the necessary ethical approvals and permissions, the process initiates. Here, in this research we had collected MRI images as the dataset that gathers neuroimaging data like MRI scans to capture structure of the brain and changes over time. This MRI images is being thrown into a web application which will analyse whether the particular individual is suffering from this disease or not based on the four kinds of images like Demented, Non- Demented, Mild Demented and Moderate Demented.

**2.3.2 FEATURE SELECTION**

Decision tree is the feature selection which is put upon. We have utilized feature engineering techniques to extract relevant features form the data, such as genetic markers, brain imaging, and cognitive assessments. Feature selection is a crucial step in the early detection of Alzheimer disease using machine learning because it helps show the most relevant and informative features from the data, which can lead to more correct and efficient models.Figure 2.6 shows the feature selection process involved in this process.This plays a major role in this research project.



**Figure 2.6** Feature Selection Process

**2.3.3 DATA PREPROCESSING**

Clean, normalize, and handle missing data to ensure the quality of the dataset. Properly preparing the data ensures that the input to your ML model is clean, consistent, and ready for analysis. First, we have collected MRI images or medical data and inspected for any inconsistencies or artifacts.Then, the data is standardized by resizing images to a common resolution and orientation, making them compatible for analysis.

Data augmentation techniques such as, rotation, and flipping were employed to increase dataset variability and enhance the model generalization. Throughout this stage, careful attention is given to support the data integrity, and any missing values are addressed through proper imputation or removal strategies. Preprocessing prepares the raw data for machine learning algorithms, improving their effectiveness in this analysis.

A diagram of a process

Description automatically generated

**Figure 2.7** Methodology

**2.3.4 CROSS-VALIDATION**

We have employed K-fold cross validation to assess model performance and to prevent overfitting. It involves systematically partitioning the dataset into multiple subsets of folds, training and evaluating the ML model multiple times. The primary goal is to assess the model's performance while ensuring it generalizes well to unseen data. In the context of Alzheimer’s detection, K-fold cross validation is often employed, where the dataset is divided into k equally sized subsets. The model is trained on k-1 folds and confirmed on the remaining fold, repeating this process k times, with each fold serving as the validation set once.

**2.3.5 EARLY DETECTION**

We have developed algorithms to show early signs of cognitive decline or risk factors for Alzheimer’s individuals. In the early detection of Alzheimer’s disease using ml, the process involves the systematic analysis of medical data, often including MRI images and clinical assessments, to find subtle and early signs of the disease.

We have used SVM (Support vector machine) algorithm which uses patient's data and gives features such as age, Demented, non-demented, very mild demented, moderate demented. Figure 2.8 defines four different kinds of Dementia. By detecting these early indicators, healthcare professionals and researchers can intervene at an earlier stage. Therefore, we will be highlighting the accuracy score, precision, and F-score values to the maximum.

A comparison of a brain scan

Description automatically generated

**Figure 2.8** Four kinds of dementia in Early detection.

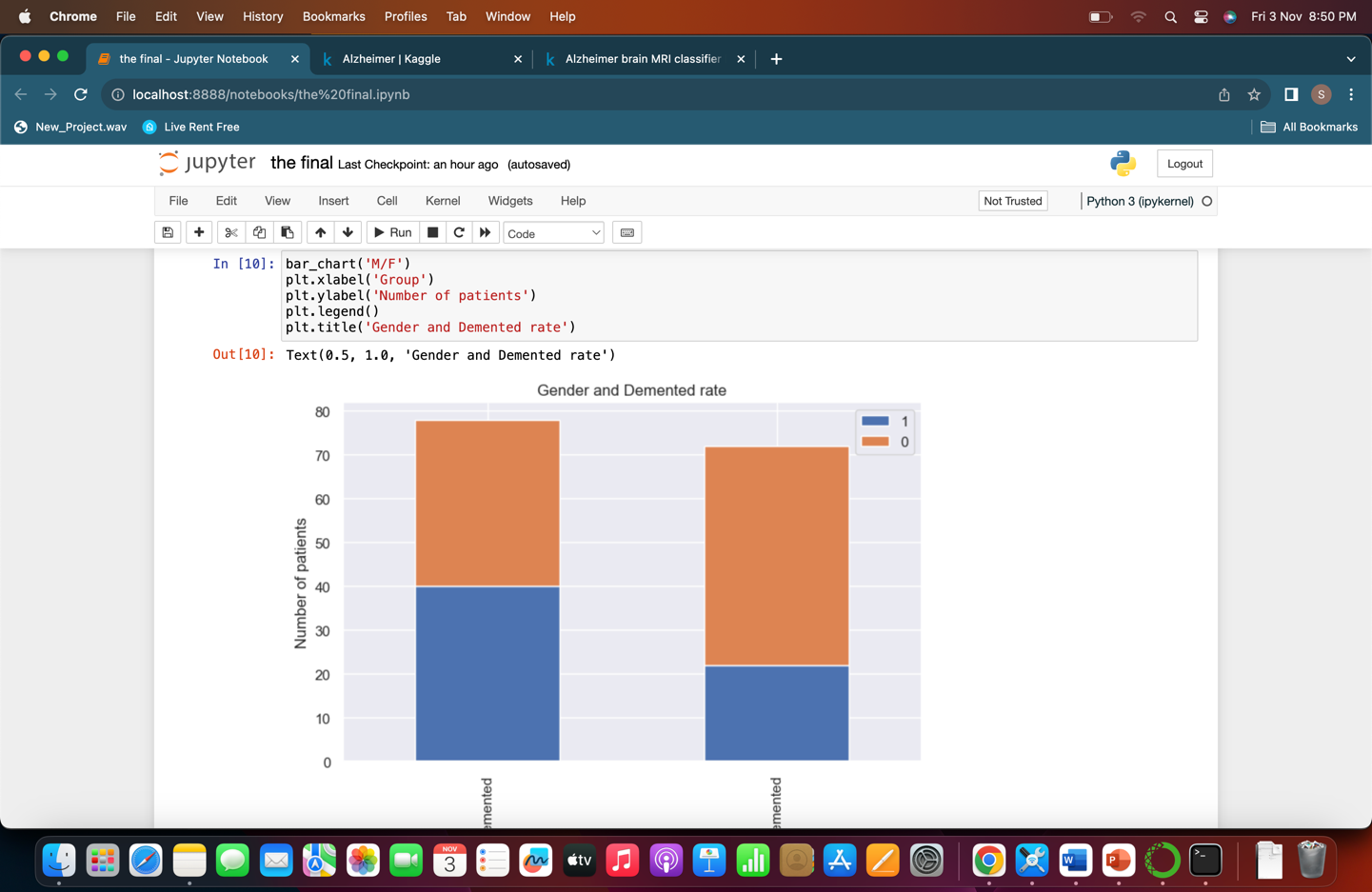
**2.4 MODULES**

The most informative features for predicting Alzheimer's disease can be identified using a variety of methods, such as statistical analysis or machine learning algorithms. Once the features have been selected, the SVM model is trained on the preprocessed data. The performance of the trained SVM model is then evaluated on a held-out test set to ensure that the model is not overfitting the training data.

Once the model is trained and evaluated, it can be deployed to a production environment to be used for the early detection of Alzheimer's disease.This can help clinicians to diagnose Alzheimer's disease earlier, when treatment is most effective. However, it is important to note that SVM-based modules are still under development and more research is needed to validate their performance in clinical settings. In addition to the potential benefits of SVM-based modules for the early detection of Alzheimer's disease, there are also some challenges that need to be addressed before these modules can be widely adopted in clinical practice.

One challenge is that SVM models can be computationally expensive to train, especially on large datasets. Another challenge is that SVM models can be sensitive to the choice of hyperparameters. It is important to carefully select the hyperparameters to optimize the performance of the model. Finally, it is important to ensure that SVM models are interpretable so that clinicians can understand the factors that are contributing to the predictions of the model.

**2.5 SCREENSHOTS AND RESULTS**

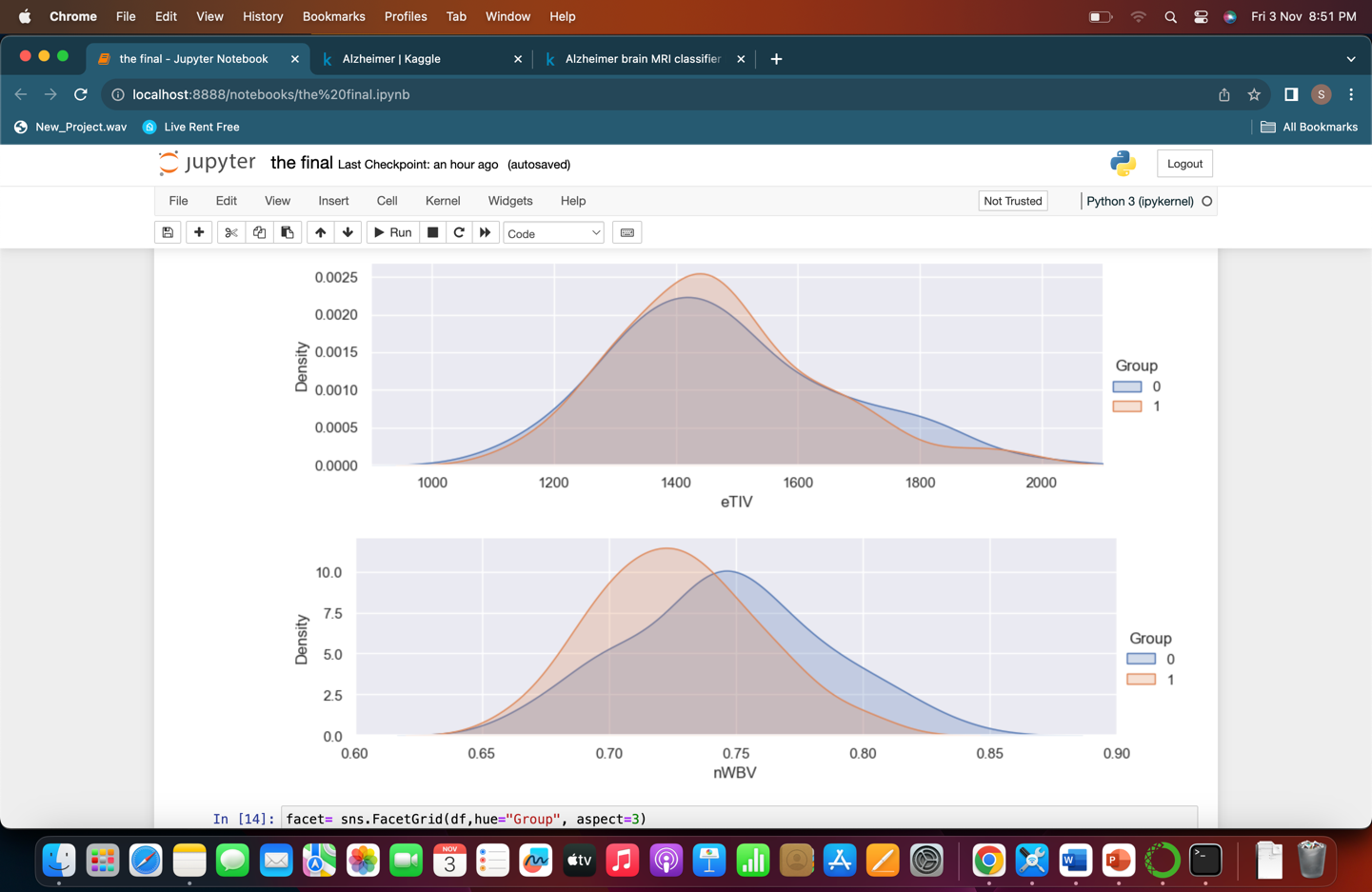


**Figure 2.9** Dementia rate for Gender

A screenshot of a computer

Description automatically generated

**Figure 2.10** Density rate of the Disease



**Figure 2.11** Comparison between dementia and Non dementia

A screen shot of a computer

Description automatically generated

**Figure 2.12** EDUC AND SES

**A graph of a disease

Description automatically generated with medium confidence**

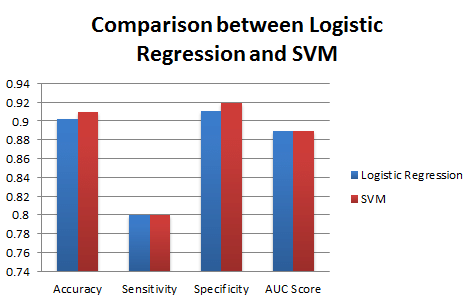
**Figure 2.13** Graph plot

**2.6 COMPARISON WITH OTHER ALGORITHMS**

In the preliminary analysis of Alzheimer’s disease using Machine learning SVM Algorithms are often compared to other machine learning algorithms.Here’s a comparison of SVM with other Algorithms and a clear explanation of the algorithms that are being compared.

**2.6.1 LOGISTIC REGRESSION**

SVMs can handle both linear and non-linear relationships between features, making them more versatile for complex datasets. Logistic regression is primarily suited for linear problems. SVMs often provide better classification performance when there is a clear margin of separation in the data.



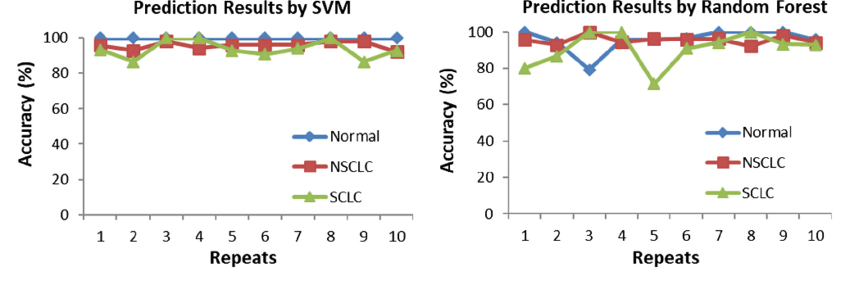
**Fig 2.14** Logistc regression and SVM

Till now , we have received an accuracy of 74.16% using logistic regression.

It is clearly seen that SVM with Random Forests-based feature selection has outperformed than Logistic Regression with RFE in terms of accuracy and specificity.The figure differentiates accuracy, sensitivity, specificity and AUC score with a colour visualization.

**2.6.2 RANDOM FOREST**

Random forests are powerful for handling high-dimensional data, while SVMs are effective in low to moderate-dimensional spaces. SVMs generally offer better interpretability, as they provide clear decision boundaries, whereas random forests generate complex decision trees that can be challenging to interpret.Figure 2.15 compares the SVM with Random forest with their performance. Till now , we have received an accuracy of 80.2% using Random forest classifier.



**Figure 2.15** Comparison between SVM and Random Forest

**2.6.3 GRADIENT BOOSTING**

SVMs provide transparent and interpretable decision boundaries, making it easier to understand how the model makes predictions. This can be particularly important in medical contexts where interpretability is crucial. Gradient Boosting algorithms, on the other hand, tend to create complex, non-linear decision boundaries. They are generally less interpretable than SVMs, but this complexity can sometimes lead to better predictive performance. Till now , we have received an accuracy of 79% using Gradient Boosting.

**2.6.4 VOTING**

SVMs can be less scalable for large datasets due to the quadratic optimization problem they solve, leading to longer training times. The Voting algorithm's scalability depends on the base models used. It can be scalable if the base models are chosen to be efficient and well-suited for the dataset. Till now , we have received an accuracy of 73.29% using Voting algorithm. However, they can struggle with highly imbalanced or noisy datasets. On the other hand, The Voting algorithm can enhance predictive accuracy by aggregating the predictions of multiple models, providing more robust results in various scenarios, including imbalanced or noisy data. In summary, the Support Vector Machine (SVM) algorithm is a powerful and versatile tool in the field of machine learning. What makes SVM unique is its ability to find the best possible line or boundary for tasks like classification or regression, even when dealing with non-linear data, thanks to the use of kernel functions. This makes SVM a popular choice for a wide range of applications. It's also well-suited for handling data with many dimensions and is built on a strong theoretical foundation, setting it apart from other algorithms.

However, it's important to acknowledge the trade-offs. SVMs can be computationally demanding, especially with large datasets, and they require careful tuning of certain parameters to perform at their best. Figure 2.16 shows the overall comparison of graph between all the algorithms that we have discussed above.Despite these challenges, when used thoughtfully, SVMs prove highly effective in a variety of real-world scenarios, making them a valuable asset in the machine learning toolbox.

A graph of blue bars

Description automatically generated

**Figure 2.16** Comparison between Algorithm

**CHAPTER 3**

**CONCLUSION AND FUTURE WORKS**

The early assessment of Alzheimer's disease using a Support Vector Machine (SVM) model in machine learning shows promising results and offers valuable insights into the potential for early detection of this debilitating condition.

**3.1 CONCLUSIONS**

The SVM model demonstrates its effectiveness in identifying patterns and markers associated with Alzheimer's disease, which can aid in its timely diagnosis. Future work in this area should focus on expanding the dataset to improve the model's accuracy and generalisability. Additionally, incorporating other machine learning techniques and exploring multimodal data sources, such as neuroimaging and genetic information, may further enhance the early assessment capabilities. Furthermore, the integration of real-time patient data and the development of user-friendly diagnostic tools can contribute to more accessible and efficient Alzheimer's disease detection, ultimately improving patient care and quality of life. There are several other critical aspects to consider. First, the incorporation of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), can enhance the model's ability to detect subtle patterns in large-scale medical data. Moreover, the development of explainable AI methods is vital to provide insights into the SVM model's decision-making process, ensuring transparency and trust in the diagnostic results. Collaborations with healthcare institutions and the collection of diverse, well-annotated datasets will be essential to improve the model's robustness and real-world applicability. Furthermore, the assessment of the model's performance against various demographic and ethnic groups should be a priority to address potential bias and disparities in diagnosis. Finally, the integration of telemedicine and wearable technology can facilitate continuous monitoring, enabling early detection in a more patient-centric approach, and thus, improving the overall management of Alzheimer's disease. The table of differentiating all the algorithm’s accuracy is given in Table 3.1 below.

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-score** |
| --- | --- | --- | --- | --- |
| Logistic regression | 75.46% | 0.80 | 0.79 | 0.78 |
| Random forest classifier | 86.92% | 0.85 | 0.81 | 0.80 |
| Support vector machine | 90.67% | 0.98 | 0.97 | 0.90 |
| Gradient Boost | 58.92% | 0.65 | 0.65 | 0.65 |
| Voting classifier | 34.12% | 0.43 | 0.43 | 0.45 |

**Table 3.1** Comparison between algorithms

**3.1 FUTURE WORKS**

In the field of Alzheimer's disease research, there are exciting possibilities on the horizon for improving the early detection of the disease using machine learning, particularly Support Vector Machine (SVM) models. Scientists will explore a wide range of medical data, including genetic, brain imaging, and clinical information, to create more accurate and efficient algorithms. These algorithms, powered by SVM models, will be crucial in identifying Alzheimer's disease at an early stage, enabling timely treatment and personalized care plans. Furthermore, future work will focus on making these machine learning models easier to understand, so that healthcare professionals and patients can trust and make sense of the diagnostic results. Collaboration between data scientists, medical experts, and technology developers will likely increase, driving innovation and expanding the possibilities for early Alzheimer's disease diagnosis. These efforts are leading us towards a future where machine learning will significantly improve the lives of individuals at risk of or affected by Alzheimer's disease.

**REFERENCES**

1. Sivakani GA, Ansari R.(2020)  “Machine learning framework for implementing Alzheimer's disease*.*” Ty - book,au - rs, sivakani,au - ansari, gufran,py - 2020/08/06.
2. Khan P, Kader MF, Islam SR, Rahman AB, Kamal MS, Toha MU, et al..(2021) “Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances. IEEE Access.” Jul 24;22(15):7911.
3. Prajapati R, Khatri U, Kwon GR. (2021)“An efficient deep neural network binary classifier for alzheimer's disease classification.” pages=231-234.
4. Khalaf OI, Abdulsahib GM, Sabbar BM.(2020) “Optimization of wireless sensor network coverage using the Bee Algorithm. J Inf Sci Eng*.*” 22, 2867.
5. Kavitha C, Mani V, Srividhya SR, Khalaf OI, Tavera Romero CA.( 2022) “Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. Front Public Health.”
6. Saratxaga CL, Moya I, Picón A, Acosta M, Moreno-Fernandez-de-Leceta A, Garrote E, et al.(2021) “MRI Deep learning-based solution forAlzheimer's Disease Prediction. J.Pers. Med.” 11(9):902. PMID: 34575679; PMCID: PMC8466762.
7. Williams, Jennifer A, Weakly A, Cook MS, Edgecombe DJ.(2018) “Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia,” PY - 2013/01/01.
8. Khalaf OI, Sabbar BM(2018). “A modified algorithm for improving lifetime WSN*.*J Eng Appl Sci.”PMID: 35309200; PMCID: PMC8927715.
9. Khalaf OI, Abdulsahib GM, Sabbar BM.(2020) “Optimization of wireless sensor network coverage using the Bee Algorithm. J Inf Sci Eng*.*” 22,2867.
10. Sudharsan M, Thailambal G (2021). “Alzheimer's disease prediction using machine learning techniques and principal component analysis.” PY - 2021/04/01.

**ONLINE REFERENCE**

1. <https://www.geeksforgeeks.org/support-vector-machine-algorithm/>
2. <https://www.sciencedirect.com/science/article/pii/S0730725X21000138>
3. <https://www.nature.com/articles/s41598-022-10202-2>
4. <https://pubs.acs.org/doi/abs/10.1021/acs.bioconjchem.9b00695>
5. <https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/>
6. <https://towardsdatascience.com/machine-learning-basics-support-vector-machine-svm-classification-205ecd28a09d>
7. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5576415/>
8. <https://onlinelibrary.wiley.com/doi/abs/10.1111/jon.12266>
9. <https://link.springer.com/article/10.14283/jpad.2022.50>
10. <https://www.javatpoint.com/machine-learning-random-forest-algorithm>