|  |
| --- |
| **TABLE OF CONTENTS** |
| **TOPIC PAGENO**  **ACKNOWLEDGEMENT** |
| **ABSTRACT** |
| **LIST OF FIGURES** |
| **LIST OF TABLES**  **LIST OF ABBREVIATIONS**  **1. INTRODUCTION** |

Introduction to [Alzheimer’s disease 1](#_bookmark0)0

[Introduction to Machine Learning 1](#_bookmark1)3

Classification of ML 14

Features of ML 15

Advantages of ML 16

[Introduction to MATLAB 17](#_bookmark2)

[Features of Matlab 17](#_bookmark3)

## [LITERATURE REVIEW 18](#_bookmark4)

## [METHODOLOGY 28](#_bookmark5)

[Dataset 28](#_bookmark6)

Dataset Description 29

[MMSE 29](#_bookmark7)

[CDR 30](#_bookmark8)

[NWBV… 31](#_bookmark9)

[EDUCATION 31](#_bookmark10)

[ASF 32](#_bookmark11)

v

Training Data 32

Test Data 32

Validation 32

Splitting data 33

Machine Learning Algorithms 34

Decision Tree Algorithms 34

Support vector machine Algorithms 36

Naïve Bayes Algorithms 37

[Performance Measures 39](#_bookmark12)

1. [SOFTWARE TOOL 42](#_bookmark13)

[Introduction to Matlab 42](#_bookmark14)

[Features of Matlab 42](#_bookmark15)

[Advantages 43](#_bookmark16)

[Platform Independence 43](#_bookmark17)

[Predefined functions 44](#_bookmark18)

[Device Independent Plotting 44](#_bookmark19)

[Graphical user Interface 44](#_bookmark20)

[Matlab Compiler 44](#_bookmark21)

[Matlab Downloading 45](#_bookmark22)

[Matlab Installation 46](#_bookmark23)

Matlab Workspace 48

[Matlab commands for managing a session 49](#_bookmark24)

Commands for working with system 49

1. [RESULTS AND DISCUSSIONS 50](#_bookmark25)
2. [CONCLUSION AND FUTURE SCOPE 59](#_bookmark26)
3. REFERENCES 60

vi

## LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **FIGURE NO** | **TITLE** | **PAGE NO** |
| 3.1 | Block Diagram of Implementation | 28 |
| 3.2 | Splitting of data | 33 |
| 3.3 | Classification using Decision Tree | 35 |
| 3.4 | SVM algorithm | 36 |
| 5.1 | Confusion matrix ofDecision tree of train data | 51 |
| 5.2 | Confusion matrix of Decision tree of test data | 52 |
| 5.3 | ROC curve od Decision Tree classifier | 53 |
| 5.4 | Confusion matrix ofSVM of train data | 54 |
| 5.5 | Confusion matrix ofSVM of test data | 55 |
| 5.6 | Confusion matrix of Naïve Bayes oftrain data | 56 |
| 5.7 | Confusion matrix of Naïve Bayes of test data | 57 |

vii

|  |  |  |
| --- | --- | --- |
|  | **LIST OF TABLES** |  |
| **TABLE NO** | **TITLE** | **PAGE NO** |
| 3.1 | MMSE score for Dementia | 30 |
| 3.2 | CDR rating for Dementia | 31 |
| 5.1 | Accuracycomparison for various classifiers | 50 |
| 5.2 | Parameters comparison for various classifiers | 58 |

|  |  |
| --- | --- |
| **ABBREVIATIONS** | **TITLE** |

|  |  |
| --- | --- |
| AD | Alzheimer’s Disease |
| ASF | Atlas Scale factor |
| CDR EDUC MMSE NWBV MRI MCI ML  SVM  SES | Clinical Dementia Rating Years of Education  Mini mental state Examination Normalize whole Brain volume Magnetic Resource Imaging Mild cognitive Impairment Machine Learning  Support Vector machine  Socio Economic status |

# INTRODUCTION

## INTRODUCTION TO ALZEIMER’S DISEASE

Alzheimer's disease (AD) is a neurological condition that typically develops gradually and gets worse over time. It is the root cause of 60–70% of dementia cases. The most prevalent initial sign is trouble recalling recent events. Language difficulties, disorientation (including a propensity to get lost), mood swings, loss of desire, self-neglect, and behavioral problems can all be indicators of advanced Alzheimer's disease. As a person's health deteriorates, they frequently isolate themselves from friends and relatives. Body functions gradually deteriorate, which eventually results in death. The typical life expectancy following diagnosis is three to nine years, though the rate of progression can vary. The progression of cognitive and functional deterioration is often defined as occurring in three stages during the course of Alzheimer's. Early or mild, middle or moderate, and late or severe are terms used to define the three stages. The early signs of memory loss are caused by the disease's recognized targeting of the hippocampus, which is connected to memory. The severity of memory loss increases as the illness worsens. In real life, the clinician chooses the tests that are best for each patient. A diagnosis is made if the findings are unmistakable. If not, additional tests are requested by the doctor for clarification. All of these choices are made with the patient in mind. Instead, most machine learning algorithms presume that all biomarkers are instantly accessible and apply the same classification model to all patients without adjusting the diagnostic choices. Huge volumes of data may now be recorded thanks to recent technological advancements. To assist in the understanding of such data for clinical diagnosis and decision-making, machine learning techniques have been proposed. However, the majority of machine learning systems now fall short of replicating the individualized diagnosis procedure found in actual clinical settings. In real life, the clinician chooses the tests that are best for each patient. A diagnosis is made if the findings are unmistakable. If not, additional tests are requested by the doctor for clarification. All of these choices are made with the patient in mind. Instead, most machine learning algorithms presume that all biomarkers are instantly accessible and apply the same classification model to all patients without adjusting the diagnostic choices.

Short-term memory loss, paranoia, and delusional thoughts are symptoms of Alzheimer's Disease (AD), a degenerative neurological disorder that is often misdiagnosed as stress or aging-related symptoms. About 5.1 million people in the US are afflicted by this illness. AD does not receive adequate medical care. AD must be treated with medication consistently. Since AD (1) is a chronic condition, you may live with it for a long time. Therefore, in order to prevent significant brain damage, it is crucial to prescribe medication at the right time. Since we need to collect a lot of data, apply advanced algorithms for prediction, and include an experienced doctor, early detection of this disease is a time-consuming and expensive process. Genetic and environmental factors can both contribute to Alzheimer's disease development. Over time, these have an impact on an individual's brain. This disease will inevitably affect someone due to genetic alterations. Over time, this illness gradually destroys brain tissue. People over 65 years old experience it. However, the average person with this condition lives with it for 9 years, and 1 in 8 individuals over 65 have it.

The primary factor used to predict the disorder is the MMSE (Mini Mental State Examination) score. If the person is impacted, this score is periodically reduced. Dementia is a severe risk for MCI patients to experience. Dementia is anticipated to emerge in the scenario when the fundamental MCI causes memory loss. The elderly are the main group of people affected by Alzheimer's Disease (AD) a neurodegenerative condition. The fact that there is no cure to slow down or stop the disease's progression is distressing because it is a progressive condition. The reports from 2005 through 2030 show a consistent increase in the percentage estimate of the number of AD sufferers. There are around 40 million AD patients globally. An intriguing aspect of AD is that, despite being incurable, early diagnosis and effective treatment can stop the destruction of neurons. Currently, Computer-Aided Diagnostics use sophisticated computer programmers and algorithms in the area of image analysis. Our brains change as we get older, just like the rest of our body. Most people eventually become aware of some thinking slow down and sporadic memory issues. Serious memory loss, disorientation, and other significant changes in how our minds function, however, might be an indication that brain cells are deteriorating.

Changes brought on by Alzheimer's disease frequently start in the area of the brain that controls learning. Insidious confusionabout events, time, and place; unfounded suspicions about family, friends, and professional caregivers; more severe memory loss and behavioral changes; and difficulty speaking, swallowing, and walking are just a few of the symptoms that become more severe as Alzheimer's spreads through the brain. Having trouble remembering people's names when you meet strangers is a common issue in the early stages of Alzheimer's disease. Other issues include forgetting a passage you just read in a book or an alternative, having trouble locating or misplacing a pricey item, and finding it increasingly difficult to complete tasks and activities.

According to a World Health Organization survey from 2022, about 10 million new instances of Alzheimer's disease are reported each year, affecting an estimated 55 million people globally. Early diagnosis of this illness requires extensive data collection, the use of sophisticated algorithms for prediction, and the participation of a knowledgeable medical professional. Automated systems can beused in medical decision support systems and are more accurate than human evaluation since they arenot subject to human error. The diagnosis of Alzheimer's will take less time if it is automated, and it will also require less human interaction, which is crucial. Additionally, automation yields more accurate outputs while lowering overall costs. It might be difficult to make a clinical diagnosis of Alzheimer's disease, especially early on. We seek to improve diagnosis efforts by utilizing classification tools. This investigation examines various techniques for classifying individuals with Alzheimer's disease based on their MRI scans and demographic data. Some of the MRI biomarkers employed in this investigation came from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database and UC Berkeley Biomarkers, respectively. The Alzheimer's Disease Neuroimaging Initiative(ADNI), a multisite study, aims to improve Alzheimer's disease clinical trials by improving their efficiency (AD) in various ways depending on the situation. Using resources fromthe public and private sectors, this collaborative study investigates people with AD, those who could develop AD, and controls who don't show anysigns ofcognitive impairment.

## INTRODUCTION TO MACHINE LEARNING

Machine learning is a technology which enables computers to learn automatically from past data. Machine learning uses a variety of techniques to create mathematical models and make predictions based on previous information or data. Machine Learning plays a vital role in health care system. Digital diagnosis is one example of how machine learning is used in the healthcare industry. Within patient electronic medical records, ML can identify patterns associated with specific diseases and alert clinicians to any irregularities. There are several algorithms in Machine learning to detect the Alzheimer’s disease at an early stage. Most of the researches has been done on every algorithm with different approaches to get the suitable algorithm(method) which gives more accuracy to detect the disease. When a machine learning system receives new data, it forecasts the outcome using the prediction models it has built using prior data. The amount of data used determines how well the output is anticipated, as a larger data set makes it easier to create a model that predicts the outcome more precisely. In this study we compared the accuracies of different algorithms and will suggest themost accurate method. The algorithms are Decision Tree, SVM [Support Vector Machine], Naïve bayes algorithms. The complete process has been done in MATLAB software. You may learn about machine learning and a variety of machine learning approaches, including supervised, unsupervised, and reinforcement learning, in this video. Regression and classification models, clustering techniques, hidden Markov models, and other sequential models will all be covered.

Machine learning is significant because it aids in the development of new goods and provides businesses with a picture of trends in consumer behaviors and operational business patterns. A significant portion of the operations of many of today's top businesses, like Facebook, Google, and Uber,revolve around machine learning. For many businesses, machine learning has emerged as a keycompetitive differentiation. One of the most well-known branches of artificial intelligence is machinelearning. Nearly every industry, including healthcare, finance, infrastructure, marketing, self-drivingcars, recommendation systems, chatbots, social media, gaming, cyber security, and many others, usesmachine learning techniques. Machine learning is currently in the development stage, and numerous new technologies are constantly being introduced. It benefits us in a variety of ways, including large-scale data analysis, data extraction, interpretation, etc.

## CLASSIFICATION

To teach a machine to learn and make predictions, detect patterns, or classify data, a lot of data must be presented to it. supervised, unsupervised, and reinforcement learning are the three categories of machine learning.

## SUPERVISED LEARNING

In supervised learning, sample labelled data is given to the machine learning system as training material, and then it uses that information to predict the outcome. The system builds a model using labelled data to comprehend the datasets and learn about each one. After training and processing, the model is tested by utilizing sample data to see if it accurately predicts the desired outcome. In supervised learning, mapping input and output data is the main objective. The foundation of supervised learning is supervision, just like when a pupil is studying under a teacher's supervision. Spam filtering is a prime example ofsupervised learning.

Several commercial goals, such as sales forecasting, inventory optimization, and fraud detection, can be accomplished by supervised learning. Use cases include, for instance:

* + - * + estimating the price ofrealestate
        + determining the degreeoffraud in bank transactions
        + Identifying illness risk elements
        + assessing the riskiness ofpotential borrowers for loans
        + predicting the failure ofmechanical components in industrialequipment

## UNSUPERVISED LEARNING

Unsupervised learning is a type of learning where a computer picks up information without any human intervention. The machine is trained using a set of unlabeled, unclassified, or uncategorized data, and the algorithm is required to respond independently on that data. Unsupervised learning's objective is to reorganize the incoming data into fresh features or a collection of objects with related patterns. There is no predefined outcome in unsupervised learning. The machine searches through the vast volume of data for helpful insights. It can also be divided into two types of algorithms 1. Clustering 2. Association.

Additionally, clustering and association, which identify the rules that exist between the clusters, are common uses. Clustering builds a model that groups things together based on particular attributes. Several instances ofuse cases include

* + - * + grouping customers based ontheir buying habits
        + grouping inventories based on manufacturing and/or revenue metrics
        + identifying relationships in customer data (for example, customers who buy a specific style ofhandbag might be interested in a specific style ofshoe)

## REINFORCEMENT LEARNING

A learning agent in a reinforcement learning system receives a reward for each correct action and receives a penalty for each incorrect activity. With the help of these feedbacks, the agent automatically learns and performs better. The agent explores and engages with the environment during reinforcement learning. Anagent performs better since itsobjective is to accruethe most rewardpoints. Reinforcement learning is demonstrated by the robotic dog, which automatically learns how to move its arms. Due tothe fact that most firms lack the necessary computational power, most ML platforms do not have reinforcement learning capabilities, according to Gartner Reinforcement learning can be used in situations that can be fully simulated, are immobile, or have a lot of pertinent data. This type of machine learning is said to be simpler to utilize e when working with unlabeled data sets because it involves less management than supervised learning. This kind of machine learning is still being used in actual applications. Examples ofcertain uses are as follows:

* + - * + teaching vehicles how to park and drive themselves
        + adjusting traffic lights dynamicallyto ease congestion
        + using unprocessed video as input to teach robots how to follow rules so they can copy the behaviors theyobserve

## FEATURES OFMACHINE LEARNING

* Data is used bymachine learning to find different patterns in a dataset.
* It can automaticallyget better bylearning from previous data.
  + It is data driventechnology.
  + Data mining and machine learning are quite similar because both processes work with vast amounts ofdata.

## ADVANTAGES

* + - * Identifies trends and patterns quickly

Large amounts of data can be reviewed by machine learning, which can identify specific trends and patterns that humans might miss. For an e-commerce site like Amazon, for instance, knowing its users' browsing patterns and past purchases enables it to offer them the appropriate goods, discounts, and reminders. It makes advantage ofthe findings to show them relevant adverts.

* + - * No need for human intervention

You no longer have to supervise your project at every stage thanks to ML. Giving computers the ability to learn enables them to make predictions and enhance algorithms on their own. Anti-virus programmers are a typical illustration of this; they learn to filter new dangers as they are identified. ML is proficient at identifying spam.

* + - * Continuouslyenhancing

ML algorithms keep become more accurate and effective as they gather experience. They can consequently make wiser selections. Take the example of creating a weather forecast model. Your algorithms become faster at making more accurate predictions as your data set expands.

* + - * handling data with severaldimensions and types

In dynamic or uncertain contexts, machine learning algorithms are adept at managing data that ismultidimensional and multivariate.

* + - * Numerous Applications

You may use ML to your advantage as an e-tailer or a healthcare provider. Where it does apply, it has the potential to assist in providing clients with a far more personalized experience while also targeting the proper people.

## INTRODUCTION TO MATLAB

A fourth-generation high-level programming language and interactive environment for numerical calculation, visualization, and programming is called MATLAB. It enables matrix manipulation, function and data charting, algorithm implementation, user interface construction, interaction with programmers written in other languages, such as C, C++, Java, and FORTRAN, data analysis, algorithmdevelopment, and modeland application creation.

Although MATLAB is primarily designed for numeric computation, symbolic computation capabilities are accessible through an optional toolbox that uses the MuPAD symbolic engine. Graphical multi-domain simulation and model-based design for embedded and dynamic systems are added via an additional programmed called Simulink.

## FEATURES OF MATLAB

* It is a high-levellanguage for creating applications and doing numericalcomputations.
* Additionally, it offers a dynamic setting for exploratory, creative, and problem-solving iterations.
* It offers a significant library of mathematical operations for solving ordinary differential equations and performing linear algebra, statistics, Fourier analysis, filtering, optimization, and numerical integration.
* It offers tools for constructing personalized plots as well as built-in graphics for visualizing data.
* The programming interface for MATLAB provides programmers withtools to enhance the performance, maintainability, and quality oftheir programmers.
* It offers resources for creating programmers withunique graphicaluser interfaces.

# LITERATURE REVIEW

## Aruchamy, Srinivasan; Haridasan, Amrita; Verma, Ankit; Bhattacharjee, Partha;Nandy, Sambhu Nath; Ram Krishna Vadali, Siva (2020). *[IEEE 2020* National Conference on Emerging Trends on Sustainable Technology and Engineering Applications (NCETSTEA) - Durgapur, India (2020.2.7-2020.2.8)] 2020 National Conference on Emerging Trends on Sustainable Technology and Engineering Applications (NCETSTEA) – Alzheimer’s Disease Detection using Machine Learning Techniques in 3D MR Images.

Srinivasan Aruchamy and Amrita Haridasan proposed the method using machine learning techniques in 3D MR images for the detection of Alzheimer's Disease. In this, study the author used a different model to detect disease at an early stage by separating grey and white matter using 3D images. The people who are suffering from AD mainly affect the grey and white matter of the brain. The white and grey matter of the brain gets damaged. The 3d image database was collected from OASIS which consists of 460 subjects’ male and female 3D image which was taken as input was visualized in three directions Axial, Coronal, and sagittal. This first feature reduction step was performedto extract prominent features and after that pre- processing of data was performed. The first- order statistical features that were retrieved from the data [16] include energy, entropy, the interquartile range, kurtosis, the first-order maximum, and mean absolute deviation. First-order minimum, robust mean absolute deviation, root mean squared, skewness, total energy, uniformity, and variance. In this Four different algorithms are used to detect AD at the early stage. The algorithms used are Logistic Regression, SVM, Naive Bayes, and Ada boost classifier. These are applied to both grey and white matter separately in all three directions. The accuracy of all the algorithms was taken. The input of 460 subjects was taken for the classification and from the classification n t the accuracy logistic regression obtained was 87.8% among all ofthemthe lowest accuracyobtained was for75.3% of Naive Bayes the grey matter. The maximum accuracy obtained was 90.9% for Ada boost algorithms. The drawback of this paper is to study the grey matter and white matter separately effort should be made to studythe 3D images. So it takes more time and more effort.

## Kim, J. P., Kim, J., Park, Y. H., Park, S. N., Seo, S. W., &Seong, J.-K. (2019). Machine learning based hierarchical classification of frontotemporal dementia and Alzheimer's disease. *Neuro Image: Clinical*, *23*, 101811.

As one of the most prevalent kinds of early-onset dementia, front temporal dementia (FTD)was hierarchically categorised in this study. A subject categorization model for each individual subject would be considerably more beneficial than a group analysis. In this study, the author's objective was to utilise a machine learning-based classification method using surface-based cortical thickness data to assign each patient to a hierarchically organised diagnostic category. The individual eventually received one ofthe final clinical designations once the tree's hierarchy of approaches continued. Each participant was successfully classified into one of five clinical groups with 75.8 percent accuracy, according to the classification findings using the entire hierarchical tree. Thus developed a machine learning-based automatic classifier for the differentiating FTD clinical symptoms. Using a fully automated classifier, corticalthickness data alone could classify FTD clinical subgroups and AD with good to outstanding accuracy.

1. **Bi,X.-an,Hu,X.,Wu,H.,&Wang,Y.(2020).MultimodalDataAnalysisofAlzheimer'sdisease based on Clustering Evolutionary Random Forest. *IEEE Journal of Biomedical and HealthInformatics*, *24*(10), 2973–2983.**

In this work, they conducted multimodal data fusion research on Alzheimer's illness. Our findings have numerous ramifications for the development of computational medicine as wellas the diagnosis of Alzheimer's disease. In order to find connections between brain regions and genes, theyfirst performed correlation analysis. their technique effectively fuses data from many modalities and makes follow- up analysis easier. Second, it is advised to examine "brain region- gene pairings"and extract the properties of the discriminative fusion that set AD apart from HC. Using this approach, they identified abnormal brain regions and AD-causing genes, such as the thalamus, lingual gyrus, angular gyrus, precuneus,insulaDAB1gene, andLRP1Bgene.However, it must be emphasized that results must be confirmedwith larger datasets as well as that additional research is required to confirm these findings and "brain region-gene combos" in the future.

**[4]J.Neelavenni, M.S.Geetha Devasana (2020). Alzheimers Disease Prediction using MachineLearning Algorithms. IEEE.**

In this research authors used different machine learning algorithms to predict the Alzheimer’s disease. As input to the model, they used various psychological parameters including MMSE (Mini Mental State Examination), age, Education and the primary factor used is the MMSE. A person is diagnosed with AD if their own MMSE score reduces gradually. Using SVM and decision trees (Machine learning algorithms), they were able to estimate the AD and later they compared accuracyresults of the algorithms. After comparing the best accuracy is given by SVM (Support Vector Machine). But in this case, only one parameter—accuracy—was taken into account, and the resultantaccuracy was extremely low in comparisontoother models. 70%ofthetraining dataset is used to traineach algorithm, and 30% of the test dataset is used to test it. This model not only distinguishes between cognitive impairment but also detects the disease in the person.

## Sakshi Singh, Komal Gaikwad (2020) Shallow learning and Deep learning techniques to detect the Alzheimer’s Disease.

In this paper authors used Deep Learning and Shallow Learning techniques to predict the Alzheimer’s disease. With the aid of various shallow learning techniques, it is also possible to study the psychological and socioeconomic effects that the disease may have on those who are affected in addition to clinical insights into the identification of Alzheimer's Disease. Machine learning techniques known as "shallow learning" reach a performance ceiling when you add more examples and training data to then network. In this studythey employed the Mann-Whitney test and correlation matrix with the Seabon library, which runs the standard uni private pearson’s test, for appropriate characteristics. They had tested the dataset by using Gradient boosting classifier, XG boost, RFC, Ada boost classifier, Decision tree classifier, SVM Linear, SVM Radial, Logistic regression. According to this study, linear SVM and Ada boost classifier closely trail logistic regression in terms of accuracy when based on clinical parameters related to brain volume and the MMSE.

## Kishore, P., Usha Kumari, C., Kumar, M. N. V. S. S., & Pavani, T. (2021). Detection andanalysis of Alzheimer’s disease using various machine learning algorithms. *Materials Today:Proceedings*, *45*, 1502–1508

As stated by the author The proposed structure shows a big processing model from a data mining standpoint. The data is provided to the system as an input. The dis organised Electronic Health Record (EHR) / Patient Health Record, clinical systems, and outside sources (PHR). The effort that went into creating an Alzheimer's rate and quality framework utilising a varietyof machine learning techniques is described in this study. The use of a categorization system makes it simpler to determinewhat stage they are in, allowing therapy to start. In order to boost the effectiveness of diagnosing thecondition inquestion, this study presents a variety of algorithms for classifying data. It is found that the Support Vector Machine with linear kernel modeloffers greater accuracy than other models**.**

## Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & L

**T. (2020). Machine learning techniques for the diagnosis of Alzheimer’s disease. ACM**

Transactions onMultimedia Computing, Communications, and Applications, 16(1s), 1–35.

165 papers fromthe years 2005 to 2019 were studied using a varietyof feature extraction and machine learning techniques. SVM, ANN, and DL, three significant machine learning algorithms for diagnosing Alzheimer's disease, were investigated by the authors. They have also researched additional learning techniques like transfer, ensemble learning, and multi-kernel learning. Alzheimer's disease has been successfully treated using SVM-based models, proving itseffectiveness. This is because SVM does not suffer from the drawbacks of local minima, unlike methods like ANN. SVM is used frequently because it is more understandable than deep neural networks, which are opaque models. Future work on this problem should focus on the clinical interpretability of deep learning models. Additionally, it has been noted that researchers have paid more attention to the feature extraction stage than the categorization stage.

1. **Liu, M., Zhang, J., Nie, D., & Shen, D. (2018). Anatomical landmark based deep feature representation for MR images in brain disease diagnosis. *IEEE Journal of Biomedical and Health Informatics*, *22*(5), 1476–1485.**

The author presented a convolutional neural network (CNN) for patch-based deep feature learning after using data to identify discriminative anatomical landmarks in MR images. They proposed methodology for disease prediction using anatomical landmark-based deep feature learning (LDFL). The proposed LDFL method has been validated in MR information extraction and brain disease classification tasks. Our research made use ofthe Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) and the Alzheimer's Disease Neuroimaging Initiative (ADNI) databases. Four main components make up the overall structure of our suggested technique. the following key elements: 1) landmark discovery, 2) landmark-based patch extraction, 3) patch-based feature learning, and 4) applications for disease categorization and image retrieval. In this article, they proposed a deep feature learning method based on landmarks for automatically deriving patch- based representations from MR images for the identification of brain illnesses associated with Alzheimer's disease. This technique paves the way for discriminative biomarkers in morphological analysis of MR images andcomputer-aided diagnosis of Alzheimer's disease.

1. **Naik, B., Mehta, A., & Shah, M. (2020). Denouements of Machine Learning and Multimodal Diagnostic Classification of Alzheimer’s disease. *Visual Computing for Industry, Biomedicine, and Art*, *3*(1).**

Chronic memory loss and cognitive decline are hallmarks of the neurological disorder known as Alzheimer's disease (AD). Numerous experiments were conducted to evaluate the effectiveness of utilising a multi classifier since its main objective is to provide the optimal hyperplane that separates data points of one class from data points of another class that were not generated. This study's main objective was to assess the precision of several ML classification techniques when used to classify AD using MRI data. Adoctor must make a quick decision regarding the best course oftreatment when Alzheimer's disease is in its advanced stages. Thus, delaying a clinical choice would be pointless and maybe dangerous for patients. The addition of the PET/SPECT/CSF modality to MRI increased classification accuracyover MRI alone, which helped the SVM classifier's performance rates. The classifier's accuracy increases when a single modality is combined with numerous modalities.

1. **Mofrad, S. A., & Lundervold, A. S. (2021). A predictive framework based on brain volume trajectories enabling early detection of Alzheimer's disease. *Computerized Medical Imaging and Graphics*, *90*, 101910.**

The objective is to identify the transition between CN, MCI, and AD (Alzheimer's Detection). The experiment's data set is ADNI. Using Free Surfer v.6.0, TW1 pictures are used to extract subcortical segments and cortical parcellation, which are then used to select the 3D regions that have a propensity for AD and fit into a model for prediction. Their mixed strategy is further broken down into I feature selection, model building, and validation, and model evaluation (ii). Predictive performance analysis uses datasets from the ADNI and the ADNI with AIBL combinations. For CN and MCI, the accuracy is roughly 73% and 78%. With the same methodologies as its base reference model, which had previously only achieved 64% efficiency among 224 candidates, the primary goal of this research is to increase accuracy and performance. Additionally, it has suggested algorithmic instability that results in poor model quality. Although the use of Free Surfer v.6.0 can be held responsible for some of the variation's reduction, instability still exists.

## Ghoraani, B., Boettcher, L. N., Hssayeni, M. D., Rosenfeld, A., Tolea, M. I., & Galvin,

**J.E. (2021). Detection of mild cognitive impairment and alzheimer’s disease using dual- task gait assessments and machine learning. *Biomedical Signal Processing and Control*, *64*, 102249.**

In this essay, the author set out to develop a method for precisely and consistently detecting MCI andAD. The goal of this work was to build an automated and objective method for detecting a cognitivedecline in MCI and AD patients using just gait analysis, as well as to identify crucial gait factors for machine learning-based categorization. The use of gait as a cognitive impairment screen may promptmedical professionals to schedule additional testing to identify MCI and AD. They recorded the gaitsof 78 elderly individuals as they walked in a variety of single- and dual-task environments. The 108 gait features of each participant were extracted, and the important uncorrelated features werefound.. The technique produces 25 significant uncorrelated gait factors for differentiatingbetween healthy and MCI, healthy and AD, and MCI and AD, as well as 13 for MCI and AD. The five-fold

classification accuracy was 78 percent using the given gait parameters, which was a little under 83 percent.

1. **Bi, X.-an, Hu, X., Wu, H., & Wang, Y. (2020). Multimodal Data Analysis of Alzheimer's disease based on Clustering Evolutionary Random Forest. *IEEE Journal of Biomedical and Health Informatics*, *24*(10), 2973–2983.**

They investigated Alzheimer's disease using multimodal data fusion in their paper. Our findings have numerous ramifications for the development of computational medicine as well as the diagnosis of Alzheimer's disease. They started by using correlation analysis to hunt for connections between particular brain regions and genes. their method efficiently combines data from many modalities and facilitates follow-up easier analysis Second, it is advised to use the CERF to assess "brain region- gene pairings" and extract the distinctive traits that set AD apart from HC. An AD diagnostic framework that also comprises feature selection, feature synthesis, and sample categorization also in corporates the CERF as a final component. Using this framework, they identified abnormal brain regions and AD-causing genes, such as the insula DAB1 gene and LRP1B gene, as well as the thalamus, lingual gyrus, angular gyrus, precuneus, etc. However, it is important to emphasise that results must be confirmed with larger data sets, and that future research must confirm these findings and "brain region-gene combos."

1. **Hazarika, R. A., Abraham, A., Sur, S. N., Maji, A. K., & Kandar, D. (2021). Different techniques for Alzheimer’s disease classification using brain images: A study. *International Journal of Multimedia Information Retrieval*, *10*(4), 199–218.**

The author claims that because routine diagnosis is costly and occasionally inaccurate, researchers are working to develop a dependable and affordable method for classifying AD using brain imaging. The result comparison demonstrates that the ANN-based categorization technique yields the most compelling conclusions(approximately 93.19 percent). Obtaining adequate data points from diverse data sources is one of the main challengesthat researchers could face (both online and offline). It is challenging to separate the alterations because of the intricate anatomy of the brain. As a result, efficient pre- processing procedures like skull stripping and component segmentation are important yet challenging.

## Kruthika, K. R., Rajeswari, M., & Maheshappa, H. D. (2019). Multistage classifier- based approach for Alzheimer's disease prediction and retrieval. Informatics in Medicine Unlocked,14, 34–42.

Dementia is a confirmed result of Alzheimer's disease (AD), which has no known cure as of this writing. Earlier stage discovery allows for more effective therapy. A better and more effective set ofdiagnostic techniques for AD was suggested in a paper, in addition to physical clinical diagnosis and neurological evaluation. The application of content-based image retrieval(CBIR) systems, which combine automated image categorization algorithms and radiologist expertise, increases prediction accuracy. The author of this paper utilised the Naive Bayes classifier, the Support Vector Machine (SVM), and Using K-nearest neighbour (KNN) to categorise the disease's stages. In this study, the clinical progression of the brain structural change associated with Alzheimer's disease was represented using the swarm intelligence technique PSO for feature selection**.** The multistage classifier used in this thesis outperformed earlier standalone machine learning methods like SVM andKNN in terms of AD detection. Additionally, the image retrieval strategy that was used after the suggested method for classifying AD produced successfulresults.

1. **Khan, N. M., Abraham, N., & Hon, M. (2019). Transfer learning with intelligent trainingdata selection for prediction of Alzheimer’s disease. *IEEE Access*, *7*, 72726– 72735.** The detection of Alzheimer's disease (AD) utilising MRI results that contain neuroimaging data was covered in this work by the author. They use transfer learning to tackle these issues, pre-training the cutting-edge VGG architecture with weights from sizable benchmark datasets of real-world photos. In this article, we suggest a transfer learning-based technique for MRI- based Alzheimer's diagnosis. Theyassert that using transfer learning withwell chosentraining data and a strong, tested architecture for natural images might not only increase a model's accuracy but also lessen its reliance on a big amount of data. Last but not least, the author offers Class Activation Maps (CAM), which show how the suggested model concentrates on discriminative picture regions that are pertinent to neuropathology and assist the medical professional in understanding the model's decision-making process.

## Fan, Z., Xu, F., Qi, X., Li, C., & Yao, L. (2019). Classification of Alzheimer’s disease based on Brain MRI and machine learning. *Neural Computing and Applications*, *32*(7), 1927–1936.

In this research, the author addressed how Alzheimer's disease (AD) affects 50–60% of individualsand is the most prevalent neurodegenerative illness, with a concealed and gradual course of progression. Memory, cognitive ability, and behaviour changes are all possible effects of Alzheimer'sdisease. According to this study, the way of categorising diseases using the information included in photographs may be more beneficial for clinical diagnosis. One of the extraction techniques that is currently most frequently employed is principal component analysis (PCA). The facial recognition technology was made possible using principal component analysis. Additionally, principal component analysis was integrated with support vector machines (SVM), linear discriminant analysis (LDA), and linear discriminant analysis (PCA) in this study to identify AD in MRI imaging data. The basic ideas, categorization methods, and learning techniques of machine learning are introduced in this work. The fundamental concept and principle of the support vector machine are also introduced at this time. It demonstrates that the SVM method may be used to assess unknown samples and has a strong abilityto suit the disease course of Alzheimer's.

## Kishore, P., Usha Kumari, C., Kumar, M. N. V. S. S., & Pavani, T. (2021). Detection and analysis of alzheimer’s disease using various machine learning algorithms. *Materials Today: Proceedings*, *45*, 1502–1508.

As stated by the author The proposed structure shows a big processing model from a data mining standpoint. The data is provided to the system as an input. The disorganised Electronic Health Record (EHR) / Patient Health Record, clinical systems, and outside sources (PHR). The effort that went into creating an Alzheimer's rate and quality framework utilising a variety of machine learning techniques is described in this study. The use of a categorization system makes it simpler to determine what stage they are in, allowing therapy to start. In order to boost the effectiveness of diagnosing the condition in question, this study presents a variety of algorithms for classifying data. It is found that the Support Vector Machine with linear kernel model offers greater accuracythan other models.

1. **Albright, J. (2019). Forecasting the progression of alzheimer's disease using neural networks and a novel preprocessing algorithm. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *5*(1), 483–491.**

In this study, machine learning approaches were investigated for their potential to estimate Alzheimer's disease development in the future using clinical data. Numerous machine learning techniques applied in this study were effective in foretelling the progression of Alzheimer's disease in both cognitively well and MCI individuals. Before someone may be diagnosed with MCI or dementia, these techniques may be able to identify those who are at high risk for Alzheimer's disease. These techniques may increase the likelihood of discovering an AD therapy because the inability to identify patients at an early stage is one of the primary causes of the recurrent failure of AD clinical trials. This study investigated the use of clinical data to estimate the future course of Alzheimer's disease using machine learning techniques.

## Acharya, U. R., Fernandes, S. L., WeiKoh, J. E., Ciaccio, E. J., Fabell, M. K., Tanik,U. J., Rajinikanth, V., & Yeong, C. H. (2019). Automated detection of alzheimer’s disease using brain MRI images– a study with various feature extraction techniques. *Journal of Medical Systems*, *43*(9).

The author of this work wants to develop a Computer-Aided-Brain-Diagnosis (CABD) system that can determine whether a brain scan shows signs of Alzheimer's disease. The quantitative approaches employed in the paradigm include filtering, feature extraction, Student's t-test based feature selection, and KNN based classification. Additionally, the current feature-based procedures are carefully examined, and feature extraction is done utilising the student's t-test. A reduced number of attributes are needed for the ST + KNN compared to previous methods. The proposed paradigm fared better than both the brain MRI data from the medical clinic and the benchmark AD dataset. Future research will substitute SVM, neural networks, random forests, and AdaBoost for KNN for classifying data.

# METHODOLOGY

With a lot of data, it becomes difficult to manipulate the data to get the desired results; therefore, it becomes necessary to design a model that is enforced by an algorithm. The flow chart of each algorithm is shown below.

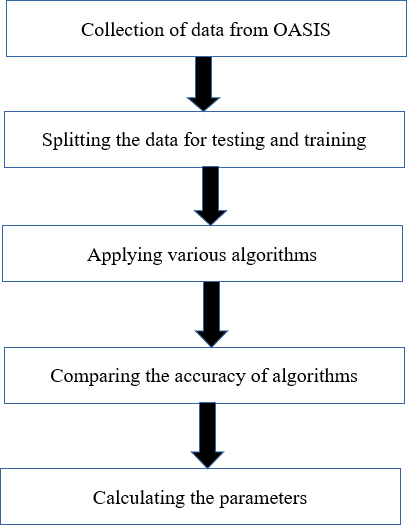


Fig 3.1 Block diagram of implementation

## DATASET

A dataset is a group of connected, discrete pieces of connected data that can be viewed separately, together, or controlled as a single unit. The input dataset is collected from the OASIS which consists of Age, Gender, MMSE, CDR, eTIV, nWBV, Hand, ASF. Thedataset contains 373 subjects aged from 60to 96. The data used was statisticaldata for predicting the Alzheimer’s disease.

## DATA DESCRIPTION

For machine learning algorithms, we have used numeric dataset which consist offeatures likeSES Socioeconomic Status

MMSE Mini Mental State Examination

ASF Atlas Scaling Factor

ETIV Estimated TotalIntracranial Volume

NWBV Normalize Whole Brain volume

CDR Clinicaldementia rating

* Education: (1 < high school(HS), 2 = HS Graduate, 3 = Some College, 4 = CollegeGraduate,5 = Beyond College Graduate)
* Socioeconomic Status (SES): (1 = lower, 2 = medium, 3 = intermediate, 4 = higher middle, 5= upper)
* Mental state examination (MMSE): (0,30). The MMSE is a 30-item survey that has beenproven to be accurate and reliable in detecting dementia.
* Totalintracranial volume (ETIV) estimate: (1132–1992) mm3. The ETIV variable measuresthe volume of the intracranial brain.
* Normalized whole brain volume (NWBV): (0.64–0.90) mg. This variable represents thetotal volume ofthe brain.
* Atlas scaling factor (ASF): (0.88–1.56) The ASF is a one-parameter scaling factor that allows comparisons of projected total intracranial volume based on human anatomy variances.

## MMSE

MMSE means Mini-Mental Status Exam. The Mini-Mental State Exam (MMSE), also known as the Standardized Mini-Mental State Examination (SMMSE), is the most used method for determining the extent of Alzheimer's disease or a similar dementia in a patient. The MMSE was created to

useful clinical assessment of how older patients' cognitive status has changed. It covers the subject's sense of place and time, memory recall, short-term memory, and mathematical process. A patient takes the MMSE, which consists of a sequence of questions meant to measure a variety of common cognitive abilities. The MMSE allows for a maximum score of 30. A person with Alzheimer's loses two to four points on the MMSE exam on average per year. This test is very helpful to detect the Alzheimer’s at an early stage. Because it is very simple to take and no specialdevices are required.

|  |  |
| --- | --- |
| **MMSE Score** | **Level of dementia** |
| 24 and higher | Normalcognition; no dementia |
| 19-23 | Mild dementia |
| 10-18 | Moderate dementia |
| 9 and lower | Severe dementia |

Table 3.1 MMSE score for Dementia

## CDR

CDR means clinical dementia rating. It was a global rating method used to identify the Alzheimer’s in the patients at an early stage. The informant interview and the subject interview yield data that is used to calculate the CDR score. Memory, Orientation, Judgment and Problem-Solving, and Personal Care are the six important factors that will helpful to calculate the overall CDR score. There are two sets of questions, one for the informant and one for the subject. The informant's questionnaire includes questions concerning the subject's memory issues, and problem-solving skills, as well as questions about the subject's involvement in local politics.

Based on the given answers we will get to know whether he/she affected with Alzheimer’s or not. The different scaling level and their symptom is shown in the below table.

|  |  |
| --- | --- |
| **CDR Rating** | **Symptom** |
| 0 | Nondementia |
| 0.5 | Verymild |
| 1 | Mild |
| 2 | Moderate |
| 3 | Severe |

Table 3.2 CDR rating for Dementia

## NWBV

Normalized whole brain volume (nwbv) reflecting the percentage of the intracranial cavity occupied by brain, was obtained using previously established methods. It is different for both male and female. If the volume is small compared to actual volume, then we can say the person is affecting from Alzheimer’s disease.

* + 1. **EDUCATION**

Education is one of related symptom to find Alzheimer’s disease in a person. But mostly it doesn’t helpful to get the accurate results. People who pursue lifelong learning are less likely to develop Alzheimer's disease, because if we want to learn something new our brain will tries to work sharply. A high level of education is supposed to boost a person’s cognitive reserve. However, not all research found an association between lower education and a higher incidence of dementia. The amount of schooling linked to dementia risk varied by research population, and more years of education did not always reduce dementia risk.

## ASF

The Atlas Scaling Factor (ASF) was computed as the determinant of the affine transform connecting each individual to the atlas-representative template. The Atlas Scaling Factor (ASF) was computed as the determinant of the affine transform connecting each individual to the atlas-representative template.

# Training Dataset

The training data is the biggest (in -size) subset of the original dataset, which is used to train or fit the machine learning model. Firstly, the training data is fed to the ML algorithms, which lets them learn how to make predictions for the given task. For supervised learning, the training data contains labels in order to train the model and make predictions. The model's accuracy and predictive power are heavily influenced by the kind of training data we give it. It implies that the model will perform with better the quality of the training data. A typical ML project's training data makes up more than or equalto 60% ofthe total data.

# Test Dataset

Once the model has been trained using the training dataset, it is time to test it using the test dataset. This dataset assesses the model's performance and guarantees that it can generalize well to new or unexplored datasets. The test dataset is a different subset ofthe original data fromthe training dataset. When the model training is finished, it utilizes it as a benchmark because it has some similarfeatures and a similar class probability distribution. A well-organized dataset called test data provides information for each type of scenario the model might encounter in the actual world. Typically, the test dataset makes up 20–25% of the overall original dataset. During the testing and training stages, we can also access and contrast the correctness of our model using the test dataset in comparison to the training dataset. The model is considered to have over fitted its accuracy on training data is higher than its accuracy on testing data.

## VALIDATION

The validation set is regarded as a component ofthe model's training and is used to fine-tune its hyper parameters. The data is just used by the model for evaluation purposes; it is not used to learn anything new about the model. When the loss of the validation dataset exceeds the loss of the training dataset, the model's training can be interrupted and used for regression instead. That is, minimizing bias and variance. To put it another way, using a big validation set will produce better results if the model has a lot of hyper parameters. Now, whenever the model's validity data accuracy. It is divided into two types. They are Cross validation and hold out validation.

# Need of splitting of the data:

Splitting the dataset into train and test sets is one of the important parts of data pre-processing, as by doing so, we can improve the performance of our model and hence give better predictability. We can think of it as if we trained our model with a training dataset and then tested it with a test dataset that is entirely different from the training dataset, at which point our model would be unable to recognize the relationships between the features. Datasets used to train and test machines to learn.

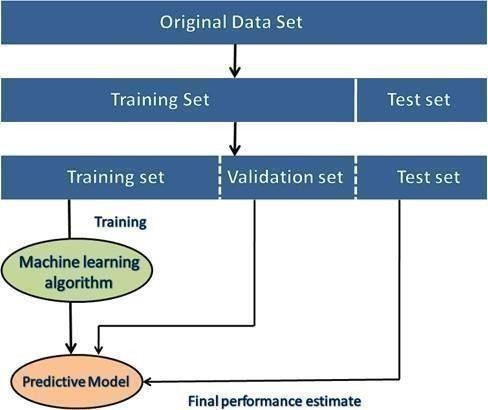


Fig 3.2 Splitting of Data

As a result, the model's performance will suffer if it is trained and tested on two different datasets. Thus, it is crucial to divide a dataset into a train set and a test set. The below flow chart shows the phases of machine learning.

* 1. **MACHINE LEARNING CLASSIFIERS**

## DECISION TREE ALGORITHM

To detect the Alzheimer’s disease at an early stage that one of best and proposed method is Decision tree algorithm. It is one of the examples of Supervised machine learning technique. Supervised learning means It is distinguished by the way it trains computers to accurately classify data or predict outcomes using labelled datasets. The training data that is given to the computers in the supervised learning serves as the instructor, instructing them on how to correctly predict the output. It employs the same idea that a pupil would learn under a teacher's guidance. The decision tree algorithm can be used for both classification and regression. As these two are types of supervised learning. The name itself indicates that it is having a tree like structure. In general, how the tree is going the same structure is following in the algorithm. In this algorithm, where internal nodes stand in for a dataset's features, branches for the decision- making process, and each leafnode for the classification result. It contains two nodes that is decision and leaf node as shown in thebelow figure. A decision is made using a Decision node, which has multiple branches, whereas a Leafnode is the result ofthat decision and does not have any additional branches.

This algorithm is very helpful in predicting the Alzheimer’s disease at an early stage. It is very easy to apply and gives accurate results also. Because Decision trees are typically designed to resemble how people think when making decisions, making them simple to comprehend. The below flow chart shows the structure of decision tree algorithm. As it contains different nodes like root node, leaf node etc. Root node means that the decisions start from this node only and the full dataset is represented, which is then split into two or more homogeneous sets. After obtaining a leaf node, the tree cannot be further divided because leaf nodes are the last output nodes.

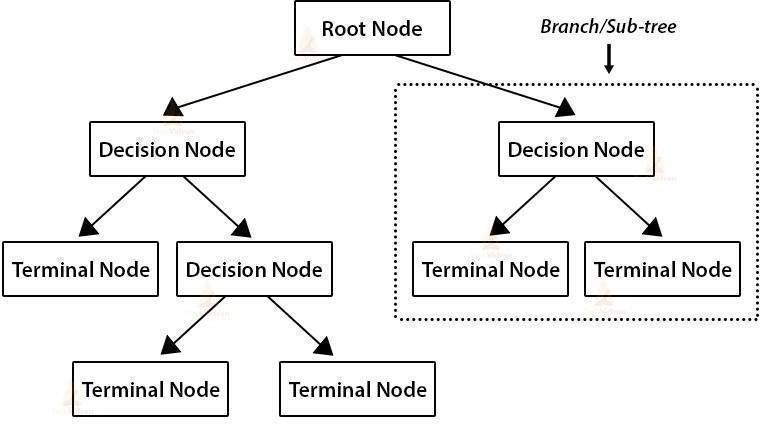


Fig 3.3 Classification using Decision Tree

Here we are implanting the decision tree algorithm in MATLAB software. So, theimplementation is as follows:

* + - * Data Pre-processing step
      * Fitting a Decision-Tree algorithmto the Training set
      * Predicting the test result
      * Test accuracy ofthe result (Creation ofConfusion matrix)
      * Visualizing the test set result.

## ADVANTAGES

1. Datanormalisation is not necessaryfor a decision tree.
2. Scaling ofdata is not necessary when using a decision tree.
3. Additionally, the construction of a decision tree is not significantly impacted by missing Values in the data.

## SUPPORT VECTOR MACHINE

Another proposed method to detect Alzheimer’s disease at an early stage is Support Vector Machine Algorithm in machine learning. It is also one of the algorithms in supervised learning technique. Similar to decision tree algorithm it is also used for both classification and regression problems. But mostly it prefers to classification problems. The main idea of SVM is to create a hyperplane to separate the class labels into categories, this can be done by linearly separating the labels in 2D plane (using a line) or one can increase the dimensions based on the threshold (middle value),linearity and the number of variables. Usually when dealing with an outlier, terms like bias and variance come into picture while classifying a data point in SVM. But a few misclassifications can abet theprocess;this is called a soft margin. These soft margins are also as support vectors.

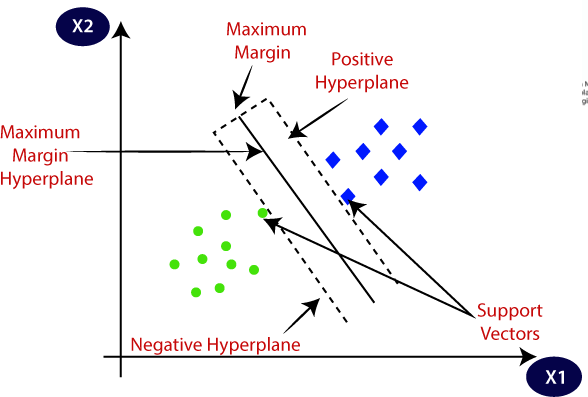


Fig 3.4 SVM Algorithm

SVM includes different planes like hyperplane. Hyper plane means in n-dimensional space, there maybe several lines or decision boundaries used to separate the classes, but we must identify the optimumdecision boundary that best aids in classifying the data points. The hyperplane of SVM is aname for this optimal boundary. The dataset's features determine the hyperplane's dimensions, therefore if thereare just two features (as in the example image), the hyperplane will be a straight line.

To implement the SVM algorithm to our dataset we have to follow certain steps. Those stepsare as given below:

* + - * Data Pre-processing step
      * Fitting a Decision-Tree algorithmto the Training set
      * Predicting the test result
      * Test accuracy ofthe result (Creation ofConfusion matrix)
      * Visualizing the test set result.

## ADVANTAGES

* + - * + Whenthere is a large gap betweenclasses, SVM performs comparativelywell.
        + In large dimensional spaces, SVM performs better.
        + Ifthere are more dimensions than samples, SVM works well in certain situations.
        + SVM uses relatively little memory.

## DISADAVANTAGES

* + - * + Large data sets are not a good fit for the SVM algorithm.
        + When the target classes are overlapping and the data set includes more noise, SVM does not perform very well.
        + The SVM will performpoorlywhen there are moretraining data samples than features for each data point.
        + There is no probabilistic justification for the classification because the support vector classifier places data points above and below the classifying hyperplane.

## NAÏVE BAYES

To predict Alzheimer’s disease, we used another algorithm in the machine learning that is Naïve bayes algorithm. As like previous algorithms it’s also a type in the supervised learning technique.

It is usedfor classification problems. This algorithm completely depends on the bayes theorem. It is mostly employed in text categorization with a large training set. It makes predictions based on object probabilities because it is a probabilistic classifier. Naïve bayes it’s a combination of two different words and there is a separate relation between these two words that is the reason for Naive is because it presumes that the occurrence ofone trait is unrelated to the occurrence of other features, it is known as naive. A red, spherical, sweet fruit, for instance, is recognised as an apple if the fruit is identified based on its color, form, and flavour. So, without relying on one another, each characteristic helps torecognize it as an apple. And the reason to called as Bayes is because it relies on the Bayes' Theoremconcept, it is known as the Bayes.

The Bayes theorem, commonly referred to as Bayes' Rule or Bayes' law, is used to calculate the likelihood of a hypothesis given some prior information. The conditional probability determinesthis. The Bayes theorem's formula is as follows:

P(A/B) = P(B/A) P(A)/P(B)

P(B|A) stands for Likelihood Probability, which measures how likely it is based onthe evidence at hand that a given hypothesis is correct.

The posterior probability, or P(A|B), measures the likelihood that a given hypothesis (A) will really occur.

To implement the naïve bayes algorithm there are certain steps involved in it. The following are the steps:

**Step 1:** Create frequencytables from the provided dataset.

**Step 2:** Create a likelihood table by calculating the oddsof the given attributes.

**Step 3:** Now, determine the posterior probability using the Bayes theorem.

## ADVANTAGES

1. Easy to Implement Evaluation of the conditionalprobability is simple.
2. It could produce excellent results if the conditional Independence assumption is true.

## DISADVANTAGES

* Naive Bayes makes the uncommon but unfounded assumption that all predictors (or features) are independent. This restricts the algorithm's usability in practicalusage cases.
* This approach encounters the "zero-frequency problem," where it gives a categorical variable with zero probability if its category was not present in the training dataset but was present in the test data set. To solve this problem, it would be better if you employed a smoothing method.
* You shouldn't take its probability outputs too seriously because its estimations may occasionally be incorrect.

## PERFORMANCE MEASURES

An evaluation ofthe classifier's performance typically involves utilizing a confusion matrix. The True classes and classifier predicted classes are presented in a specific table. In the confusion matrix, four different terminologies are used. False positives and False negatives are present, as well as True positives and True negatives [11]. Further accuracy and parameters were calculated for the classification model. With the help of a confusion matrix, these parameters are calculated for each algorithm.

## Confusion matrix

The easiest way of checking the performance of an algorithm is by generating the confusion matrix. Itis done by creating a table of four components. True positive, False Positive, True Negative and FalseNegative.

Actual

1 0

1

|  |  |
| --- | --- |
| True positive | False Positive |
| False negative | True Negative |

predicted

0

1. **True positives (TP)** – When the output of the actual class and the predicted class is 1, it iscalled a true positive.
2. **True Negatives (TN)** - When the output of the actual class and the predicted class is0, itisa true negative.
3. **False Positive (FP)** – When the value of actual class is 0 and the predicted class is 1, it is calledFalse Positive.
4. **False Negative (FN)** – when the value of actual class is 1 and the predicted class is 0, it is calledFalse negative.

## Accuracy:

From the confusion matrix, the accuracy ofthe algorithmcan be determined.

It can be defined as the ration of number ofcorrect predictions to the total number of predictions. Accuracy= *TP+TN/TP+TN+FP+FN*

## Precision:

It can be defined as the number of correct documents returned bythe ML model. It is calculated usingthe confusion matrix as well.

Precision = *TP/TP+FP*

## Recall or sensitivity:

It is defined as the number of positives returned by the ml model, which can be calculatedbyusing the confusion matrix.

Recall= *TP/TP+FN*

## F1 Score:

It can be calculated using the harmonic mean of precision and recall. The values thatF1 score can take is (1,0); with 1 being the best and 0 being the worst score.

F1= 2\* (precision \* recall) / (precision + recall)

## Specificity:

In contrast to the sensitivity, specificity is the number of negatives returned by the ml models. Specificity = *TN/TN+FP*

# 4. SOFTWARETOOL

## MATLAB

A software tool called MATLAB provides a programming environment, high-performance mathematical computation, and visualisation. With hundreds of built-in features for technical computing, graphics, and animations, it offers an interactive environment. MATLAB stands for Matrix Laboratory. MATLAB was written initially to implement a simple approach to matrix software developed by the LINPACK (Linear system package) and EISPACK (Eigen system package) projects. With its sophisticated data structures, integrated editing and debugging tools, and support for object- oriented programming, MATLAB is a contemporary programming language environment. It is multi- paradigm, MATLAB. Therefore, it can be used with a variety of programming paradigms, including functional, Object-Oriented, and Visual. Given that the word "Matrix" appears in its name, MATLAB bases all of its computations on mathematical matrices and arrays. All MATLAB variables, whether of the integer, character, or string types, only store data in the form of an array. MATLAB is capable of a wide range of operations, including matrix manipulation, algorithm implementation, data and function graphing, as well as interacting with otherprogramming languages' programmes. The chores of declaring the variables' data type and allocating an adequate quantity of storage for the variables are handled bythe MATLAB environment.

## FEATURES OFMATLAB

1. It is intended for bothsymbolic and numericalcomputing.
2. The principal applications ofthis high-level language are in engineering and scientific computing.
3. The desktop environment in which it operates offers all the capabilities needed for iterative exploration, design, and problem-solving**.**
4. Modern algorithms are used by the majority of these functions. There are many of these for both animations and 2-D and 3-D graphics.
5. To execute those programmes from within MATLAB, the application supports an external interface. The user can create his own functions in the MATLAB language, thus he is not restricted to using the built-in functions.
6. Customplots can be made using the built-in Graphics to visualise data and tools.
7. Any particular sort of problem can be addressed by specific applications, such as data classification, controlsystemdesign and tuning, and signal analysis.
8. It builds a variety of engineering, scientific, and custom user interface applications by offering anumber ofadd-ontoolboxes.
9. Additionalextra "toolboxes" are also offered by MATLAB's creators. These function collections, or toolboxes, were created for common uses such symbolic computations, image processing, statistics, controlsystemdesign, and neural networks.
10. The matrix is one of MATLAB's primary building blocks. The array is the most basic form of data. Real matrices, scalars, complex matrices, vectors, and scalars are all automatically handledas variations on the fundamental data type. Matrix data and matrix functions are MATLAB's favourite. Vector functions are optimised for by the built-in functions. Vectorized commands or programmes therefore execute much more quickly in MATLAB.

## ADVANTAGES

There are several advantages ofMATLAB programming language

* + 1. **Platform Independence**

As MATLAB runs on a variety of computers, it offers a high degree of platform freedom. Linux, various UNIX versions, Macintosh, Windows 2000/XP/Vista, and other operating systems all support the language. Applications created onanyplatform will function fullyon any other platform, and data files created for any platform can ostensibly be read on any other platform.

* + 1. **Predefined Functions**

A sizable library of predefined functions included with MATLAB offers tried-and-true answers to many common technical problems. Assume, for instance, that we are creating a software that must assess the statistics related to a collection of input data. To perform computations like the arithmetic mean, standard deviation, median, and other calculations, we would typically need to develop our own subroutines or functions in the majority of languages.

There are numerous toolboxes for specific purposes that can be used to assist in the solution of challenging issues in specific fields, in addition to the huge libraries of services built into the fundamental MATLAB language.

* + 1. **Device-Independent Plotting**

The basic plotting and imaging instructions in MATLAB are numerous. Any graphical output device offered by the machine running MATLAB can display the plots and images. This featureelevates MATLAB to the level ofa superior technicaldata visualisation tool.

* + 1. **Graphical User Interface**

A program's Graphical User Interface (GUI) can be created interactively by a programmer using a toolfound in MATLAB. With this capacity, a programmer can create sophisticated data- analysis algorithms that relatively inexperienced people can utilise.

* + 1. **MATLAB Compiler**

By converting MATLAB applications into a machine-independent p-code and then interpreting the p- code instruction at runtime, MATLAB is able to be flexible and independent of platforms. This approach is comparable to that of Microsoft's Visual Basic programming language. Unfortunately, because the MATLAB code is interpreted rather than compiled, the resulting programmes may run slowly. There is a different MATLAB compiler available.

## MATLAB DOWNLOADING

There are various MATLAB environments accessible. For instance, it is made available independently for usage by individuals, businesses, schools, and start-ups. Different flavours relateto various software designed for various purposes.

The following are the steps required for downloading

## Step1:

Visit the official website for MATLAB at [www.mathworks.com](http://www.mathworks.com/) to download it; depending on whereyou are, you could be prompted to click or routed to a page in a different language or your own country.

## Step2:

Click onthe trial software link in the tryor buy part of the webpage’s footer at the bottom.

## Step3:

A new web page loads after selecting the link for the trial software. Fill up the field with youremail address.

## Step4:

Inorder to help MathWorks build your account and give you access to their products, you are required for more information after choosing to continue. It also asks what you intend to use the MATLAB software for; select either personalise or hobbyist use, and then click Create. It also asks for the location and the email address you alreadygave.

## Step5:

A new page will now open; choose the choice to keep using your current email to proceed. Open the MathWorks email that has just been delivered to your mailbox. To do so right away, click the "Verify your email" button.

## Step6:

When the new page loads, your confirmed email address will be visible there. I agree button is located beneath it; click it to confirm, then click submit. We've signed up with MathWorks now. **Step 7:**

The next pagewill offer a selection. These are all possible options. You have the optionto move forward byselecting one ofthese options or not.

## Step 8:

It will download the MATLAB installer first, and that installer will then configure the prerequisites for the MATLAB environment. In this instance, we'll chose Windows (64-bit), but you can pick any other operating system. As a result, a pop-up window will show up on the screen when we click the Windowslink. After selecting the Save File option, a binaryexecutable file will start downloading. When you accessthe Downloads folder on your computer, you will see an icon labelled "Downloading of MATLAB" in the area where all of your downloaded items are saved. The binary executable file fort the installation is represented by this icon.

## MATLAB INSTALLATION

Step1:

Double-click the MATLAB icon. After clicking the icon, a pop-up will appear asking you to start theinstaller. Choose Run. AMathWorks Installer window will show up onthe screen.

We'll select the first option, Log in using a MathWorks Account, since it is now the default selection. Check your internet connection to ensure the MATLAB environment is setup correctly. Choose Next from the window's bottom menu as a result fromthe below figure.

After clicking yes to accept the licencing terms on the following screen, click the Next button once more.

The first option, Log in to your MathWorks Account, is automatically selected when a new page loads. Enter your MathWorks account creation email address and password here. Click next while bearing in mind the illustration below.

## Step2:

In the licencing option box that opens, a licence ID that has previously been selected will be highlighted with a blue background. Choosing your licence id here requires you to click Next once more. Your licence id is the one that was saved during STEP 9 of the installer's download (we suggested you to write it down at that time).

Simply click Next in the newly appearing Folder Selection window; changing the folder location is notnecessary for the installation ofMATLAB.

## Step3:

The following is the Product Selection window. It must be chosen because MATLAB 9.6 is the MATLAB environment. You can choose as many of the additional products as you'd like before pressing the Next button.

When the Installation Selections window opens, select the settings you want. If you decide to make a change, you can always click the Back buttonto go back to the previous stage.

After that, the Confirmation window will show up.

## Step4:

Once everything has been downloaded and installed, a message advising that MATLAB needs to been abled appears. Simply press the Next buttonnow.

Selecting Next brings up a new window that explains the activation's significance. Click Next to proceed.

Another window will pop up with your email address and the licence ID for your items displayed.

To proceed, click the Confirm button.

After the installation is finished, MATLAB and all of its associated products are installed successfully. Now press the Finish button.

## Step5:

A shortcut for MATLAB will be createdonthe desktop in accordance with the choices we set throughout the installation process. We can now use MATLAB by choosing the MATLAB Downloading icon that has been added onthe desktop.

## MATLAB Environment (Introduction to workspace)

A MATLAB shortcut will be created on the desktop according on the choices we made throughout the installation process. We can now use MATLAB by choosing the MATLAB Downloading icon that has been added onthe desktop.

The topthree windows kinds are as follows

* + Command window
  + Edit window
  + Figure window

Command window: Aplace where orders can be entered.

Edit window: This enables users to create and edit MATLAB programmes. figure window: There are plot and graphdisplays available.

Figure window:

Everygraphics command we copied in the command window is sent to the figure window or graphics, a unique grey window with a (by default) white backdrop color, to receive its result. The client can create an infinite number of figure windows ifthere is sufficient system memory.

Command window:

This is the main window. It is shown by the MATLAB command prompt (>>). When you launch thefunction programme in MATLAB, you are shown this window. All MATLAB commands,

including those for executing user-written programmes, are entered in this window. This MATLABwindow is a part ofthe programme that also has other, more compact windows or panes.

## MATLAB COMMANDS FOR MANAGING A SESSION:

1. clc: clears command window.
2. clear: Removes variable from the memory.
3. exist: Checks for existence offile or variable.
4. Global: Declares variables to be global.
5. Help: Searches for a help topic.
6. Look for: Searches help entries for a keyword.
7. Quit: Stops MATLAB.
   1. **COMMANDS FOR WORKING WITH THE SYSTEM:**
8. cd: Changes current directory.
9. date: Displays current date.
10. delete: Deletes a file.
11. diary: Switches on/offdiary file recording.
12. dir: Lists all files in current directory.
13. load: Loads workspace variables froma file.
14. Path: Displays search path.
15. Save: Saves workspace variables in a file.
16. Type: Displays contentsofa file.

# RESULTS AND DISCUSSIONS

The algorithms used in this project are **SVM, Decision Tree and Naïve Bayes, algorithms.** Among all these algorithms the highest accuracy is obtained for Decision Tree. Confusion matrix are obtained for training and testing the data using each algorithmand the accuracy is observed.

|  |  |
| --- | --- |
| **ALGORITHM** | **ACCURACY** |
| Decision Tree | 93.7% |
| Naïve Bayes | 85.1% |
| SVM | 58.1% |

Table 5.1 Accuracycomparison of various Algorithms

The Decision Tree Algorithm has more accuracy than the other algorithms. In this 80% of data is used for training and 20% of data been used for testing the classifier. Decision tree learned about the data during training, so if you now give it the same data to forecast,it will produce a prefect match. The main goalof our project is to predict the Alzhemier’sdisease at an early stage. In this project we compared three different machinelearning algorithms to find the best algorithm for our model. The dataset which is used inthis modelcontains “373” subjects with different age groups.

The confusion matrix obtained for decisiontree algorithmoftrain data is shown below fig 5.1

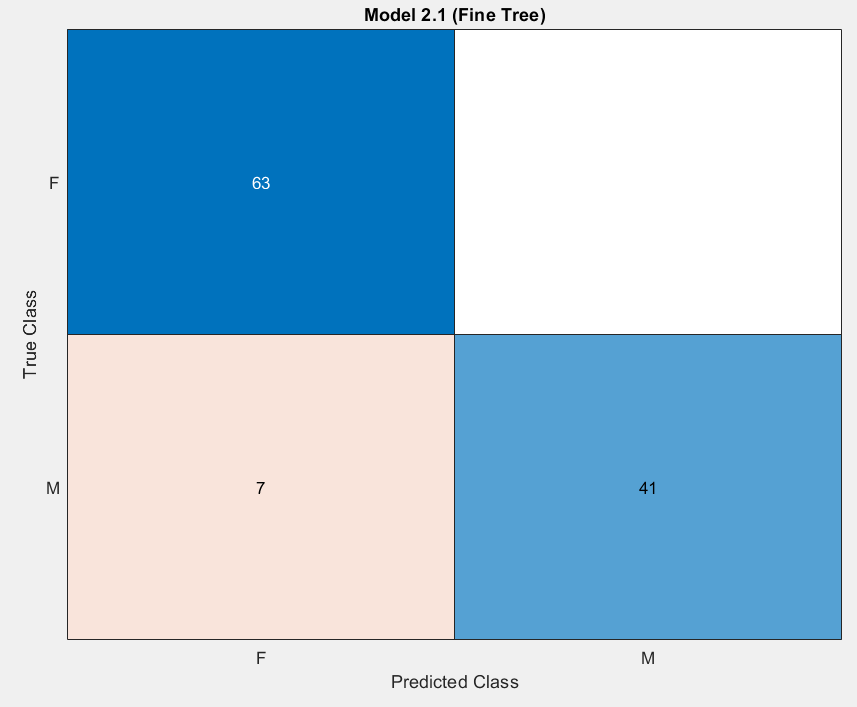


Fig 5.1 Confusion matrix of Decision Tree classifier of train data

In the confusion matrix for Decision tree algorithm the value of true positive is 63, false positive value is 0, false negative value is 7, true negative value is 41. Parameters obtainedusing confusion matrix which are as follows:

* Precision = 1
* Recall= 0.9
* F1 score = 0.94
* Specificity= 1

In Figure The 63 AD patients are correctly identified as AD patients, 7 nonAD patientsare wrongly classified as AD patients and 41 non-AD patients are correctly identified asnon-AD patients.

Department of ECE,GMRIT Page 42

The confusion matrix obtained for decisiontree algorithmoftrain data is shown below fig 5.2

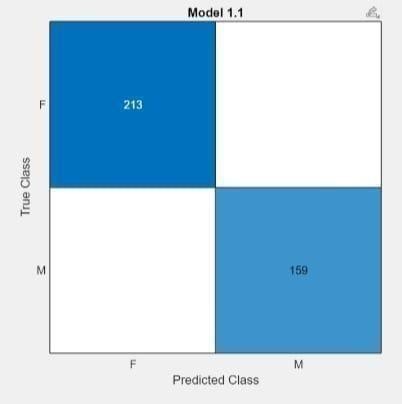


Fig 5.2 Confusion matrix of Decision Tree classifier of test data

In the confusion matrix for Decision tree algorithm the value oftrue positive is 213, falsepositive value is 0, false negative value is 0, true negative value is 169.

In Figure The 213 AD patients are correctly identified as AD patients and 159 non-ADpatientsare correctly identified as non-AD patients

## 5.1 ROC CURVE

An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plotstwo parameters:

* + True Positive Rate
  + False Positive Rate

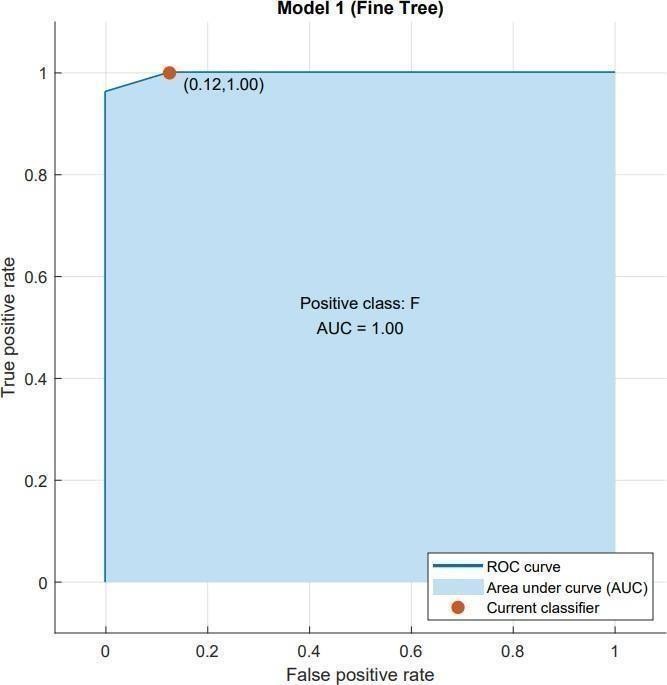


Fig 5.3 ROC curve of Decision Tree Classifier

The Figure is ROC curve shows the relationship between the true positive rate (TPR) forthe model and the false positive rate (FPR) of decision tree. The TPR is the rate at whichthe classifier predicts “positive” for observations that are “positive.” The FPR is the rate atwhich the classifier predicts

“positive” for observations that are actually “negative.”

The confusion matrix ofsupport vector machine algorithmoftrain data is shown in below 5.4

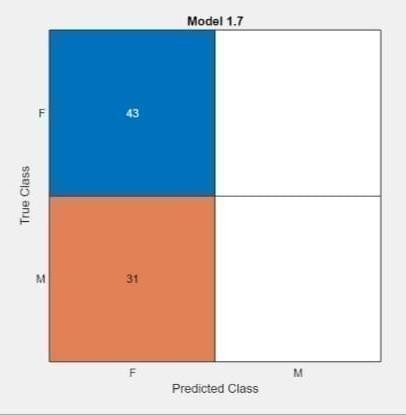


Fig 5.4 Confusion matrix of SVM classifier of train data

In the confusion matrix for Support vector machine algorithm the value of true positive is 43, false positive value is 0, false negative value is 31, true negative value is 0. Parameters obtained using confusion matrix which are as follows:

* + - Precision = 1
    - Recall= 0.58
    - F1 score = 0.73
    - Specificity= 0

In Figure The 43 AD patients are correctly identified as AD patients, 31 non ADpatients are wronglyclassified as AD patients.

The confusion matrixofsupport vector machine algorithm oftest data is shown in below fig 5.5

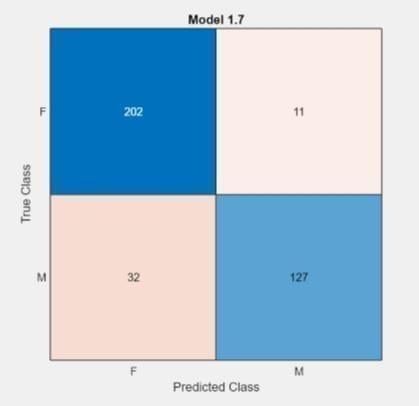


Fig 5.5 Confusion matrix of SVM classifier of test data

In the confusion matrix for Support vector machine algorithm the value of true positive is 202, falsepositive value is 11, false negative value is 32, true negative value is 127.

In Figure The 202 AD patients are correctly identified as AD patients, 11 AD patients is wrongly identified as non-AD patients, 32 non AD patients is wrongly identified as AD patients and127 non-AD patients are correctly identified as non-AD patients.

The confusion matrix ofNaïve bayes algorithmoftrain data is shown in below fig 5.6

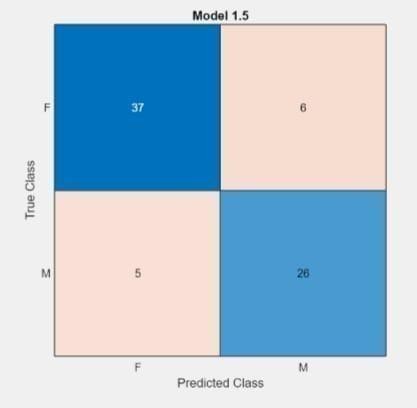


Fig 5.6 Confusion matrix of SVM classifier of train data

In the confusion matrix for Naïve Bayes algorithm the value of true positive is 37, false positive value is 6, false negative value is 5, true negative value is 26. Parameters obtained using confusion matrix which are as follows:

* + Precision = 0.86
  + Recall= 0.88
  + F1 score = 0.86
  + Specificity= 0.81

In Figure The 37 AD patients are correctly identified as AD patients , 6 AD patients is wrongly identified as non-AD patients, 5 non AD patients is wrongly identified as AD patients and26 non- AD patients are correctly identified as non-ADpatients.

The confusion matrix ofNaïve bayes algorithm oftest data is shown in below fig 5.7

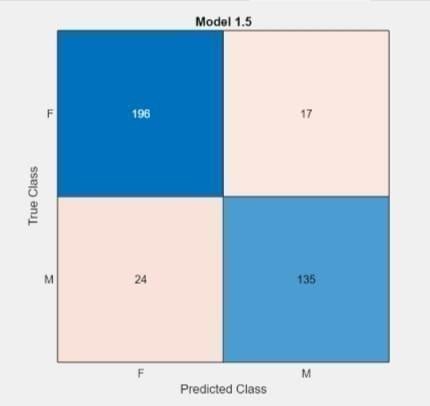


Fig 5.7 Confusion matrix of Naïve Bayes classifier of test data

In Figure The 196 AD patients are correctly identified Ad patients, 17 AD patients is wrongly identified as non-AD patients, 24 non AD patients is wrongly identified as AD patients and135 non-AD patients are correctly identified as non-AD patients.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithm** | **Precision** | **Recall** | **F1 score** | **Specificity** |
| Decision Tree | 1 | 0.9 | 0.94 | 1 |
| Naïve Bayes | 0.86 | 0.88 | 0.86 | 0.81 |
| SVM | 1 | 0.56 | 0.73 | 0 |

Table 5.2 Parameters comparison of various Algorithms

The above table shows different parameters of various machine learning algorithms. Weconsider Precision, F1- Score, Recall and Specificity. These concepts are very importantto design a prefect machine learning model which gives more accurate results. The precision of the ml model will be high when true positive value greater than thesummation of true positives and the false positives and vice versa. Recall, in contrast to precision, is unaffected by the quantity of incorrect sample classification. Additionally, Recall will be 1 if the model labels all positive data as positive.

# CONCLUSION AND FUTURE SCOPE

Alzheimer’s Disease which is also called Dementia. It is a challenging task to detect the Alzheimer’s. It does not have complete cure but predicting this disease will help to reduce the risk. Hence early prediction is more important. So, we are designing a model which will helpful to predict Alzheimer’s at an early stage. For this we are using differentmachine learning algorithms. The results obtained from this model shows that Decision tree algorithm gives best performance when compared to other machine learning algorithms. The future study will concentrate on extracting and analyzing novel featuresthat are more likely to help detect Alzheimer’s disease and removing such features. In further we can use merge different datasets which has more subjects and we can implement in “python” also.

# REFERNECES

1. Aruchamy, Srinivasan; Haridasan, Amrita; Verma, Ankit; Bhattacharjee, Partha; Nandy, Sambhu Nath;Ram Krishna Vadali, Siva (2020). *[IEEE 2020* National Conference on EmergingTrends on Sustainable Technology and Engineering Applications (NCETSTEA) - Durgapur, India (2020.2.7- 2020.2.8)] 2020 National Conference on Emerging Trends on Sustainable Technology and Engineering Applications (NCETSTEA) – Alzheimer’s Disease Detection using Machine Learning Techniques in 3D MR Images.
2. Kim, J. P., Kim, J., Park, Y. H., Park, S. N., Seo, S. W., &Seong, J.-K. (2019). Machine learning based hierarchical classification of frontotemporal dementia and Alzheimer's disease. *Neuro Image: Clinical*, *23*, 101811.
3. Bi,X.-an,Hu,X.,Wu,H.,&Wang,Y.(2020).MultimodalDataAnalysisofAlzheimer'sdisease based on Clustering Evolutionary Random Forest. *IEEE Journal of Biomedical and HealthInformatics*, *24*(10),2973–2983.

[4]J.Neelavenni, M.S.Geetha Devasana (2020). Alzheimers Disease Prediction using Machine LearningAlgorithms. IEEE.

1. Sakshi Singh, Komal Gaikwad (2020) Shallow learning and Deep learning techniques todetect the Alzheimer’s Disease.
2. Kishore, P., Usha Kumari, C., Kumar, M. N. V. S. S., & Pavani, T. (2021). Detection andanalysisof Alzheimer’s disease using various machine learning algorithms. *Materials Today:Proceedings*, *45*, 1502– 1508.
3. Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C. T. (2020).Machine learning techniques for the diagnosis of Alzheimer’s disease.
4. Liu, M., Zhang, J., Nie, D., & Shen, D. (2018). Anatomical landmark based deep feature representation for MR images in brain disease diagnosis. *IEEE Journal of Biomedical and HealthInformatics*, *22*(5), 1476–1485.
5. Mofrad, S. A., & Lundervold, A. S. (2021). A predictive framework based on brain volume trajectoriesenabling early detection of Alzheimer's disease. Computerized Medical Imaging and Graphics, 90
6. Ghoraani, B., Boettcher, L. N., Hssayeni, M. D., Rosenfeld, A., Tolea, M. I., & Galvin, J.E.(2021). Detection of mild cognitive impairment and alzheimer’s disease using dual-task gait assessments and machine learning. *Biomedical Signal Processing and Control*, *64*, 102249.
7. Bi, X.-an, Hu, X., Wu, H., & Wang, Y. (2020). Multimodal Data Analysis of Alzheimer's disease based on Clustering Evolutionary Random Forest. *IEEE Journal of Biomedical and Health Informatics*,*24*(10), 2973–2983.
8. Hazarika, R. A., Abraham, A., Sur, S. N., Maji, A. K., & Kandar, D. (2021). Different techniques for Alzheimer’s disease classification using brain images: Astudy. *International Journal of Multimedia Information Retrieval*, *10*(4), 199–218.
9. Kruthika, K. R., Rajeswari, M., & Maheshappa, H. D. (2019). Multistage classifier- based approachfor Alzheimer's disease prediction and retrieval. Informatics in Medicine Unlocked,14, 34– 42.
10. Khan, N. M., Abraham, N., & Hon, M. (2019). Transfer learning with intelligent training data selection for prediction of Alzheimer’s disease. *IEEE Access*, *7*, 72726–72735.