Predicting Alzheimer's Disease at Low Cost Using Machine Learning

By

Musfiquer Rhman	ID: 17181103098
Farjana Rahaman	ID: 16173103013
Md. Mintu Hossain	ID: 17181103110
Khadija Akter	ID: 17181103089
Umma Habiba Emu	ID: 17181103106

Submitted in partial fulfillment of the requirements of the degree of

Bachelor of Science in

Computer Science and Engineering

Under the supervisor

Dr. Muhammad Firoz Mridha

Department of Computer Science and Engineering

Bangladesh University of Business and Technology

December 12, 2021

Abstract

Modeling and predicting Alzheimer's disease (AD) is a noteworthy field of study for many years. However, diagnosing and treating AD for poor people in many developing and underdeveloped countries is a luxury. Many countries lack the proper tools. Many people do not have the capability of bearing the cost of Magnetic Resonance Imaging (MRI) tests. In this paper, we analyzed a way of identifying AD as cheaply as possible. We investigated the difference in efficiency between the Machine learning models using MRI data versus models without MRI data. We had around a 3% difference between the best models of these tests. We used the Open Access Series of Imaging Studies (OASIS) dataset. We used patients' longitudinal lifestyle data like age, gender, education, income, Mini-Mental State Exam (MMSE) score, and other features for Non-MRI models. For MRI models, we also analyzed MRI data along with the lifestyle data. We trained different Machine learning models like Random Forest Classifier, GaussianNB, LinearSVC, Logistic Regression, KNeighbors Classifier, Adaboost Classifier, and several other models. After that, we combined our best models and created a hybrid model. Our best result was 96.07% accuracy with MRI data and 93.37% accuracy without MRI data. The non-MRI result may not be as efficient as the models with the MRI dataset and not crucially inefficient. Our result can be summarized as the Non-MRI data can be used for starting the diagnosis for those poor people.

Declaration

We do hereby declare that the research work presented in this thesis entitled ***"Predicting Alzheimer's Disease at Low Cost Using Machine Learning"*** *is a result of our own work, carried out in the Department of Computer Science and Engineering, Bangladesh University of Business and Technology under the supervision of Dr. Muhammad Firoz Mridha. We also declare that it has not been submitted elsewhere for a degree, award or diploma or for any other purposes except for publications. Materials obtained from other sources are acknowledged in this thesis.*

Musfiquer Rhman	_____
ID: 17181103098	Signature
Farjana Rahaman	_____
ID: 16173103013	signature
Md. Mintu Hossain	_____
ID: 17181103110	signature
Khadija Akter	_____
ID: 17181103089	signature
Umma Habiba Emu	_____
ID: 17181103106	signature

Certification

This is to certify that we are students of B.Sc in CSE, have completed the research work titled **"Predicting Alzheimer's Disease at Low Cost Using Machine Learning"** satisfactorily in partial for the requirement of the Bachelor of Science in the Department of Computer Science and Engineering, Bangladesh University of Business and Technology in the year 2021.

Dr. Muhammad Firoz Mridha
Chairman and Associate Professor
Department of Computer Science and Engineering
Bangladesh University of Business and Technology

Dedication

Dedicated to our parents and family members for their support, love and inspirations.

Acknowledgement

First of all, we express our gratefulness to Almighty Allah, who has given us his divining blessing, patient, mental and physical strength to complete this research works.

We are deeply indebted to our project supervisor, Dr. Muhammad Firoz Mridha, Chairman and Associate Professor, Department of Computer Science and Engineering, Bangladesh University of Business and Technology. His scholarly guidance, essential suggestions, work for going through and correcting our mistakes, and generating courage from the beginning to the end of the research work made the completion of this research possible. Without his support, it would be tough for us to reach a realistic goal.

Very special gratitude goes to all our friends for their support and help to implement our work. Discussions with them on various work topics has been constructive for us to enrich our knowledge and conception regarding the work.

Finally, we are grateful to all our faculty members of the CSE department, for and making us ready for this research work with the proper guidance and supports throughout the last four years.

Approval

We acknowledge that the research works presented in this thesis entitled **Predicting Alzheimer's Disease at Low Cost Using Machine Learning** result from the original works carried out by Dr. Muhammad Firoz Mridha Chairman and Associate Professor Department of Computer Science and Engineering, Bangladesh University of Business and Technology. We further declare that no part of this thesis has been submitted elsewhere for the requirements of any degree, award or diploma, or any other purposes except for publications. We further certify that the dissertation meets the requirements and standard for the Bachelors degree in Computer Science and Engineering.

Dr. Muhammad Firoz Mridha
Chairman and Associate Professor
Department of Computer Science and Engineering
Bangladesh University of Business and Technology

List of Abbreviations

AD Alzheimer's disease

MRI Magnetic Resonance Imaging

MMSE Mini-Mental State Exam

PET Positron Emission Tomography

MCI Mild Cognitive Impairment

AUC Area Under the Curve

ROC Receiver Operating Characteristic

CDR Clinical Dementia Rating

OASIS Open Access Series of Imaging Studies

List of Tables

List of Figures

Contents

Chapter 1

Introduction

1.1 Introduction

Alzheimer's disease is a progressive neurological disorder, it causes the brain to shrink and brain cells to die. Someone is diagnosed with Alzheimer's disease every four seconds. Alzheimer's disease is one of the most common forms of dementia affecting millions of old people worldwide. Dementia refers to diseases that can be characterized by a loss of memory or other cognitive impairments. It is caused by damage to nerve cells within the brain. It becomes more severe over time and it's not curable.

The damage begins in the region of the brain which is responsible for controlling memory. The process starts several years before the first symptom. The loss of the neuron spreads to other regions and later the brain shrinks significantly. Alzheimer's disease has three stages. Mild, moderate, and severe stages. In the mild stage of Alzheimer's disease, a person can function normally but he will face problems like coming up with the right name or word, difficulty performing social tasks, forgetting things just after reading it.

The longest stage of Alzheimer's disease is the middle stage. At this stage, the person needs extra care to function properly. Symptoms at this stage are like, forgetting personal history, name, address, etc. Confusion about where they are, change in sleep pattern, etc. At the severe stage, people lose the ability

to respond to their environment, forget about their recent experience and lose awareness about their surroundings, have difficulty communicating, etc.

Every country has elderly with dementia. However, the number is alarmingly increasing in the least developing and developing nations. Research [1] has shown that by 2040, 71% of dementia patients will be from developing nations, and the number will double every 20 years. There is low awareness about dementia in those countries. Many people in those countries think the loss of memory is a typical aging problem. The lack of diagnostic assessment [2], cost of diagnosis deters people from diagnosis of dementia. The average cost of MRI in the USA is 1100 USD, the MMSE is 1.23 USD [3]. In Bangladesh, the price of a brain MRI is about 46 USD. In India, it is about 136 USD. So, Brain MRI is significantly more expensive than MMSE, which many families cannot bear.

There are several factors in dementia. Age and gender are also essential factors. Females are more like to be affected by AD than Male [4]. Low education level [5] is also a crucial factor. Low social class, income, depression, head trauma, epilepsy, diabetes, and stroke are relevant factors [6]. All these factors can be used with MMSE to early diagnosis of AD. The lower cost of diagnosis will open doors for many poor people and people living in rural areas for AD diagnosis.

Early detection of Alzheimer's disease is very helpful to start proper treatment. If the disease is predicted earlier, the progression of the symptoms of the disease can be slow down and can save lives. Alzheimer's disease can be detected early by MRI, it uses a magnetic field and radio waves to create a 3D image of the brain. Or, by analyzing statistical data and Mini-Mental State Exam. Many machine learning methods have been tested for this task in recent years, including support vector machines, independent component analysis, and penalized regression.

1.2 Problem Statement

Predicting Alzheimer's disease early has a significant effect on the treatment and minimizing the damage in patients. But most of the studies in this field rely heavily on MRI methods. Which is costly. However, non-MRI data, like longitudinal lifestyle data analysis is cheaper and can be done without the need of expensive testing Equipment. This will open doors for poor people who can not afford the expensive MRI test. However, the efficiency difference between MRI and non-MRI is not measured. In our research, We investigated the existing Machine learning algorithms to produce the best result possible for both MRI and non-MRI data, to measure the efficiency difference between the two methods. Our research is the first research to perform this benchmark.

1.3 Problem Background

AD is a common form of dementia, yet not many people in our country are aware of the disease. In underdeveloped and developing nations, many people consider it a normal aging problem. One of the main reasons behind this is the cost of diagnosis and the lack of diagnostic assessment. So a cheap way of diagnosis can be very effective in raising awareness and starting proper treatment for the affected individuals. Though MRI tests are very effective in identifying the disease, there are some other methods like longitudinal lifestyle analysis, biomarkers analysis, and cognitive tests like MMSE and Clinical Dementia Rating (CDR). Though cognitive tests alone cannot be used for the classification of a disease. However, combined with Longitudinal lifestyle data analysis can be used as a cheaper way of AD diagnosis. We are looking forward to investigating how effective this non-MRI method will be compared to the MRI test diagnosis.

1.4 Research Objectives

The objectives of our research work are:

- Prediction of Alzheimer's disease in a person.
- Minimize the cost of diagnosis for a patient while maintaining good accuracy.
- Test existing Machine learning algorithms for both MRI and non-MRI models to find out the best, more accurate architecture that performs great on the dataset.
- Create an ensemble model to maximize accuracy.
- Both MRI and non-MRI data methods will be tested and benchmarked on different matrices.

1.5 Motivations

Predicting AD is a very popular field of research for many years. A lot of progress has been done in this field. Most of the research focuses on classifying the disease perfectly. With the advancement of Machine learning and Deep learning, many researchers have found ways of identifying the disease with great accuracy [7]. But less work has been done in terms of identifying the disease cheaply so that poor people can afford the diagnose and start treatment. We investigated a cheaper way of diagnosing AD. We tested both MRI and non-MRI data and benchmarked them, to find out if the cheaper method is viable. We developed an ensemble architecture that has good accuracy and is cheaper in cost, so that poor people can afford it without compromising much accuracy. This can also be used for raising awareness about AD.

1.6 Flow of the Research

This research work is developed in several steps. First, we analyzed the research topics and studied the application of Machine Learning in this particular topic. Then we studied the existing literature. We found a lack of research in terms of low-cost diagnosis of AD. Then we investigated the existing datasets and selected one for our research. After that, we implemented several machine learning algorithms, including an Ensemble architecture, for both MRI and non-MRI data. We compared the results of both of our experiments and benchmark them on several metrics. Figure 1.1 illustrates the overall steps to the research procedure in the following diagram.

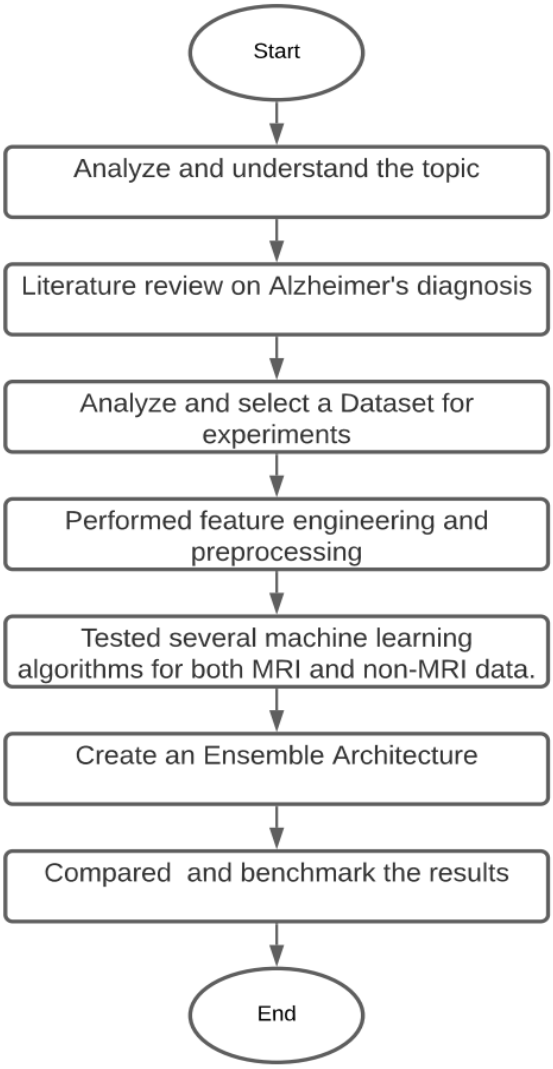

Figure 1.1: Flow of the Research

1.7 Significance of the Research

We analyzed several Machine learning models on MRI data and nor-MRI data to measure the gap between those two approaches. Analyzing MRI data gives more accurate results but is expensive. Though analyzing non-MRI data is less efficient than the MRI method. But, the efficiency difference can be drastically reduced by utilizing the effectiveness of the popular Machine learning algorithms. We proved that that analyzing non-MRI data is a viable option for starting the AD treatment. When there is a lack of diagnostic tools, or if the patient has financial problems, the non-MRI method can be used to diagnose AD. We created an ensemble architecture that can predict AD with great accuracy.

1.8 Research Contribution

Our research can be summarized as,

- We investigated the existing Machine learning algorithms to produce the best result possible for both of our tests, with and without MRI data. Our research is the first research to perform this benchmark.
- We experimented with several Machine learning algorithms and found AdaBoost gives the best result in MRI data analysis, and our Ensemble model gives the best result for non-MRI data analysis.

1.9 Thesis Organization

This thesis paper is organized as follows,

- Chapter two highlights the background and literature review about Alzheimer's disease prediction.
- Chapter three contains the AD prediction model, used algorithms, and a detailed description of the overall procedures.

- Chapter four covers the details of the tests, matrices used to evaluate our proposed architecture, and the benchmark of MRI and non-MRI data analysis.
- Chapter five explains the Standards, Impacts, Challenges, Constraints, Timeline, and Gantt Chart.
- Chapter six contains the overall conclusion of our thesis work.

1.10 Summery

This chapter includes a general overview of the problem. We looked at our research background, objectives, motivation, and the workflow of our research. This chapter also has a glimpse of what we found.

Chapter 2

Background and Related Work

2.1 Introduction

Alzheimer's disease is a popular research field for researchers. A significant amount of work has been done in this field. Researchers have worked on several areas like MRI and Positron Emission Tomography (PET) analysis, Biomarkers analysis, longitudinal data analysis, predicting patients' future progression, regional situations, reasons behind the diseases, and many more. However, more study is needed to diagnose and treat the disease at a lower cost and how to spread awareness in emerging nations.

2.2 Literature Review

Predicting Alzheimer's disease has inducted significant attention among Machine learning and Deep learning researchers. MRI and PET data analysis is the dominant subfield in predicting Alzheimer's disease. Biomarkers analysis and longitudinal data analysis for classification and progression are also much popular. A few of them are,

Gopi Battineni et al. [8] used MRI data to construct various machine learning models for predicting dementia in the elderly. Their study included

people experiencing brain abnormalities such as mild atrophy, leukoaraiosis, and regular dementia cases of Alzheimer's disease. They trained four Machine Learning models. The combination of all four models with selective features enhanced the Accuracy of dementia prediction. The combination of all four models with selective features increased the Accuracy to 98%.

Siqi Liu et al. [9] proposed a stacked auto-encoder deep learning architecture with a SoftMax output and applied it on ADNI MRI images. They could do it with less minimal domain prior knowledge and labeled training samples. They produced an overall Accuracy score of 87.76% in the classification of AD.

Yanyan Lin et al. [10] studied to develop a longitudinal structural magnetic resonance imaging-based prediction system for Mild Cognitive Impairment (MCI) progression. They collected longitudinal data from 164 MCI patients. To identify MCI patches, they used a discriminative dictionary learning framework instead of segmenting regions of interest. They had a 97% accuracy rate.

Courtney Cocherane et al. [11] analyzed different preprocessing methods, machine learning models, and feature selection techniques. Instead of using MRI data, they used longitudinal lifestyle interventions. They achieved more than 90% accuracy and recall in predicting Alzheimer's disease. They produced a "lean" diagnostic protocol that can predict AD development in someone with only three tests and four clinical visits with 87% accuracy and 79% recall.

XIN HONG et al. [12] focused on identify time relative biomarkers associated with disease status. They found that the Cortical Thickness Average (TA) is significant in predicting Alzheimer's disease progression. They propose a predicting model based on Long short-term memory (LSTM), which might connect previous information to the present task.

Garam Lee et al. [13] developed a framework that blends cross-sectional neuroimaging biomarkers at baseline, longitudinal cerebrospinal fluid (CSF), and cognitive performance biomarkers obtained from ADNI. They took advantage of the longitudinal and multi-modal nature of available data for discovering nonlinear patterns associated with MCI progression. The proposed

framework integrates longitudinal multi-domain data. The most significant advantage of their approach is that irregular longitudinal data can be used. When they used longitudinal multi-domain data, their model had an accuracy of 81 percent.

Manu Subramoniam et al. [14] proposed classifying Alzheimer's disease based on a deep neural network using Magnetic Resonance Images (MRI) as input for the classification task. They have proved that among the VGG architectures, the VGG-16 performed better than VGG-19. Among the residual neural network architectures, Resnet-18 was more accurate than Resnet-101.

Simon F. Eskildsen et al. [15] tried to predict AD using patterns of cortical thickness measurements in subjects with MCI. They were able to identify individuals with MCI who progress to AD and individuals with MCI who do not progress to AD. They also identified Specific patterns of atrophy and selected features as regions of interest from these patterns. Their prediction accuracies were artificially inflated to a range of 73% to 81%.

Anees Abrola et. al. [16] studied the progression from MCI to Alzheimer's disease (AD) by modifying deep residual neural networks. They trained the deep learning models using mild cognitive impairment individuals only, then used a domain transfer learning version and trained additionally on AD and controls. They achieved a test classification accuracy of 83.01%.

Minh Nguyena et. al. [17] showed that the minimalRNN model was better than other baseline algorithms for the longitudinal prediction of multimodal AD biomarkers and clinical diagnosis of participants up to 6 years into the future. They explored three different strategies to handle the missing data issue widespread in longitudinal data. they found that the RNN model can itself be used to fill in the missing data, thus providing an integrative strategy to handle the missing data issue., They found that after training the RNN model can perform reasonably well using one input time point with longitudinal data, suggesting the approach might also work for cross-sectional data.

Adrien Payan et. al. [18] was able to discriminate between a healthy brain

and a diseased brain by analyzing magnetic resonance imaging images as input. They used 3D convolutions on the whole MRI image which yield better performance than 2D convolutions on slices in our experiments. They also found that a 3D approach may boost the classification performance by a small margin. The accuracy was 89.47

Baiying Lei et. al [19] introduce deep and joint learning along with a two scenarios regression model for AD scores prediction. Different from the commonly used scores prediction methods which focus on machine learning or deep learning based on baseline dataset, they obtain the predicted scores at the next time point from previous time points datasets. Specifically, they integrate the feature selection with fused smoothness term and employ the correntropy to construct the joint learning model. Meanwhile, the DPN algorithm is proposed to further improve feature representation, and then SVR is applied to predict four types of clinical scores. Despite the good performance of the proposed model, there are also several limitations of their current study. The relevant clinical details like gender, age, education level, and other physiological factors of AD were not taken into account in the experiments.

2.3 Problem Analysis

Alzheimer's disease diagnosis at a lower cost needs a lot more research and analysis. These techniques only discuss diagnosing the disease as perfectly as possible using various methods. None focused on the differences in efficiency between MRI and non-MRI data analysis. In this paper, we did precisely that analysis to determine the efficiency difference. We have discussed how accurately and cheaply Alzheimer's can be diagnosed. We also proposed an ensemble model that scores great accuracy in analyzing non-MRI data.

2.4 Summary

This chapter investigated and reviewed the latest techniques of Alzheimer's disease diagnosis. The thesis's target is to create a lower-cost method of diagnosis. And benchmark the efficiency difference of MRI and non-MRI methods.

Chapter 3

Proposed Method

3.1 Introduction

This section discussed the feasibility analysis of Alzheimer's disease diagnosis at low cost and the requirements demanded in this structure. Finally, this chapter illustrates the model's overall architecture, the algorithms used and the ensemble models.

3.2 Feasibility Analysis

This research work needed five researchers with one supervisor and took nine months to develop. The thesis work required technical support in both hardware and software. This thesis work did not require any financial support from the institution or supervisor. The models were developed in a cloud environment for reducing hardware costs. To conduct the proposed architecture of the overall requirements include,

- Cloud computing service.
- Open-source software libraries for scientific computations.
- Open-source software libraries to implement the machine learning models.

3.3 Research Methodology

In this section, the methodology of the proposed architecture is portrayed. This section is sub-sectioned into three segments. The sub-sections are data preprocessing, algorithms, and ensemble architecture with detailed explanations. Figure 3.1 presents the overall work-flow of the process.

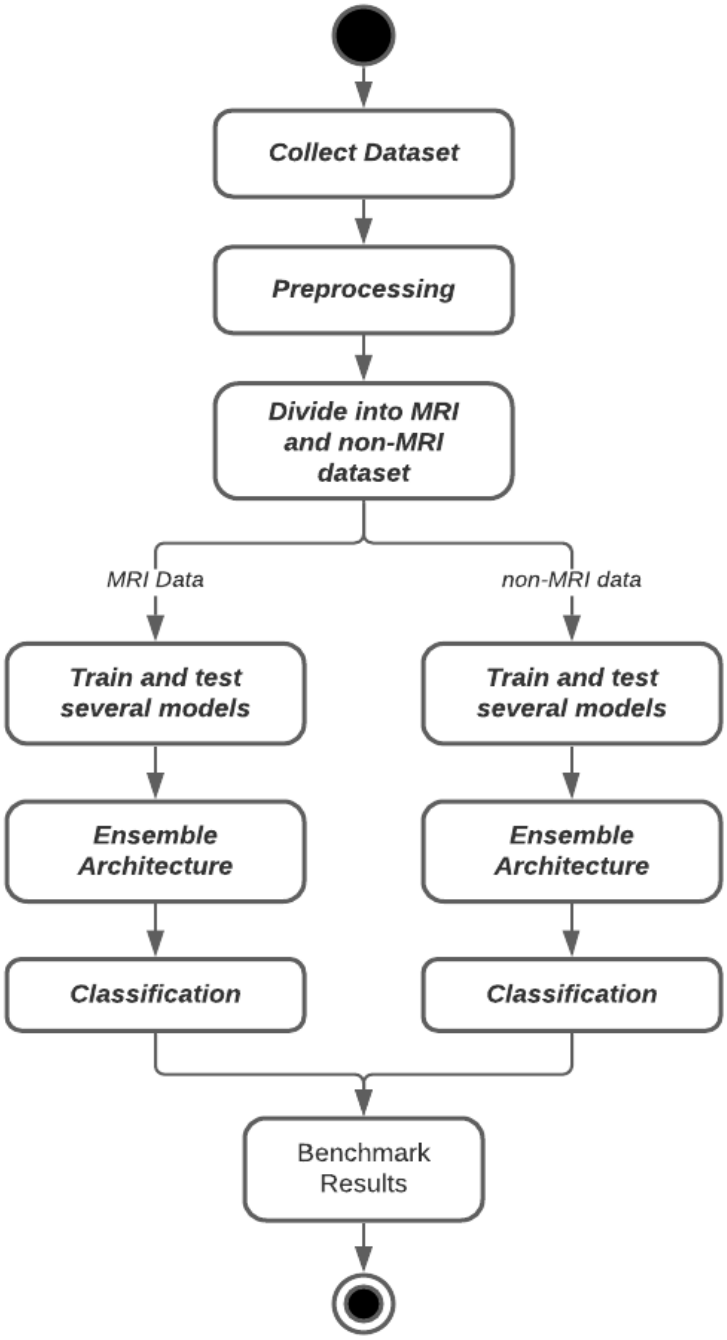

Figure 3.1: Work flow of the proposed method

3.3.1 Data Pre-processing

We used the OASIS dataset, details about this dataset is discussed in the Chapter four. This data set had some rows with empty data. So, we removed all the rows with empty data. Initially, we had 373 subjects, but after modifying our dataset, we had 354 subjects. Many machine learning algorithms cannot operate on label data directly. They require all input variables and output variables to be numeric. So, we used the One-hot encoder to encode all categorical variables and performed feature engineering. Finally, we split the data into two parts, the train set and the test set. We used 80% of the data for training our models and 20% for testing purposes. We conducted our experiment in two steps. In the first step, we included all the MRI data and longitudinal lifestyle data like age, gender, income, education, and other lifestyle data and neuropsychological scores like MMSE. In the second step, we excluded all the MRI-related data and did the same experiment again. Our dataset has multiple classes like demented, non-demented, and converters.

3.3.2 Algorithms

After processing our data, we moved onto selecting an efficient Machine learning. Selecting the best machine learning algorithm is very important. So, we experimented with several machine learning algorithms. We used both classification and ensemble learner algorithms. Ensemble learners can sometimes perform better than the classification algorithms by combining the predictions of multiple base estimators. Detailed description of the used algorithms are given below.

3.3.2.1 Random Forest Classifier

The Random Forest algorithm is a decision tree algorithm. A decision tree is consists of three nodes, decision nodes, leaf nodes, and a root node. The decision tree algorithm divides the training dataset into branches. Each branch is further

divided into other branches. This process continues until a leaf node is attained. Leaf nodes provide the classification of a given instance.

The Random forest algorithm is consists of a large number of individual decision trees. Each individual tree in the random forest makes a class prediction. The class with the most predictions becomes the model's final prediction. This process makes Random Forest more accurate than the Decision tree algorithm.

3.3.2.2 Gaussian Naive Bayes

Naive Bayes is based on the Bayes theorem. It is a simple but powerful supervised learning algorithm. This algorithm assumes that all the features of the dataset are independent. For this reason, Naive Bayes has high accuracy and speed when trained with a large dataset and the features are independent. The formula of Bayes theorem is,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A).P(B|A)}{P(B)}$$

Gaussian Naive Bayes assumes that the continuous values associated with each class are distributed according to Gaussian distribution. The likelihood of the features is,

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\alpha_y^2}} exp\left(-\frac{(x_i - \mu_y)^2}{2\alpha_y^2}\right)$$

3.3.2.3 Logistic Regression

Logistic regression is a regression analysis used for binary classification. It calculates the probability of a discrete outcome for the given input. Logistic regression is like linear regression but the range is bounded between zero and one. If the predicted probability is less than 0.5, the instance will be classified as class zero. If the predicted probability is greater than 0.5, the instance will be classified as class one. The inputs and output variables don't have to have a linear relationship, because logistic regression uses a non-linear log transformation. Logistic regression can be generalized for multiclass

classification. The logistic function is,

$$LogisticFunction = \frac{1}{1 + e^{-x}}$$

3.3.2.4 AdaBoost Classifier

AdaBoost is the short form of Adaptive Boosting. It is used together with other Machine learning algorithms to improve performance. This other algorithm is called "Weak learner". The output of the weak learner is combined with the final output as a weighted sum to boost the classifier. The AdaBoost algorithm creates decision trees by using weak classifiers. Each feature of the dataset is used to make a stump. A stump is a node with two leaves. A stump represents a feature and the leaves are the output of that tree. Each stump has a different weight based on how accurately it classifies the dataset. The weight of the incorrectly classified instances is then increased and the random samples are picked from the original dataset to create a new sample set. In this new sample set, an instance can be added several times in the new sample set. This sample is then passed into a new stump and the process is repeated for all the features.

3.3.2.5 Linear Support Vector Classifier

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms. The SVM algorithm creates the best line or decision boundary that can separate n-dimensional space into classes. This best decision boundary is called a hyperplane. Linear SVM is used for linearly separable data. Linearly separable means the dataset can be classified into two classes by using a single straight line. The data points or vectors that are the closest to the hyperplane affect the position of the hyperplane. These data points are called Support Vector. As these vectors support the hyperplane, hence the algorithm is called a Support vector.

3.3.2.6 KNeighbors Classifier

KNeighbors Classifier is a basic supervised learning classifier. It classifies an unknown data point by measuring the distance with its neighbors. Each tuple in the dataset is represented as a data point in an n-dimensional space. The KNeighbors Classifier will compare a new data point with its K existing neighbor data points, and it will classify the data point as class zero most of its neighbors are class zero. The distance is measured using the Minkowski distance. Standard practice is to take an odd value for K, to avoid any ambiguity. A low value of K can result in the underfitting of the model, A higher value of K can result in overfitting of the model. So a balanced value of K is very important. This algorithm performs poorly is the dataset is noisy.

3.3.3 Hyperparameter Tuning

Hyperparameters define the architecture of the model. Hyperparameters govern the training process. The same machine learning model can require different constraints, weights, or learning rates to generalize different data patterns. So we have to tune the parameters so that the model can optimally solve the machine learning problem. First, we defined a grid of parameters and multiple values for each parameter. Then we used GridSearchCV to find the best hyperparameters for the model. We repeated this process several times to find the best hyperparameters for our model. The supervised learning module performs a stratified ten-fold cross-validation. For example, for the Random Forest method, we tunned n_estimators, max_depth, max_features, min_samples_split, min_samples_leaf, and bootstrap for tunning the model. We created a list of different values for each and searched for the best parameters using grid search. After finding the best parameters the first time, for the numerical value parameters like nestimators, we changed the list of values, this time, we took values close to the previous best and ran again to get the maximum accuracy recall possible. We repeated this process until the Accuracy was not increasing anymore. Then using the parameters that

maximized the Accuracy and Recall, we predicted the results for our test dataset.

3.3.4 Ensemble Architecture

Ensemble Method is a machine learning technique where several Machine learning models are used to produce an optimal model. Ensemble algorithms perform better if there are significant differences or diversity among the models. There are two types of Ensemble Method, averaging methods and boosting method.

- Averaging method: Several base algorithms are built independently and then the average of the output is taken as the final predictions.
- Boosting Method: In this method base algorithms are built sequentially and one algorithm tries to minimize the bias of the combined algorithm.

After Training and tuning several machine learning models, we made an ensemble architecture. In our model, we used the Averaging method. We used the soft voting classifier to ensemble the models. The Soft voting classifier sums the probabilities of all the models and predicts the model with the highest probability. For the model with MRI data, we ensembled four models they are, Random Forest, LinearSVC, Logistic regression and Ada Boost. For The non-MRI model, we ensembled five models, the fifth model is the Gaussian NB. We didn't use KNeighbours model because it performed poorly. We tried several combinations and found these combinations yield the highest accuracy and recall score. Figure 3.2 presents a figure of an Ensemble architecture.

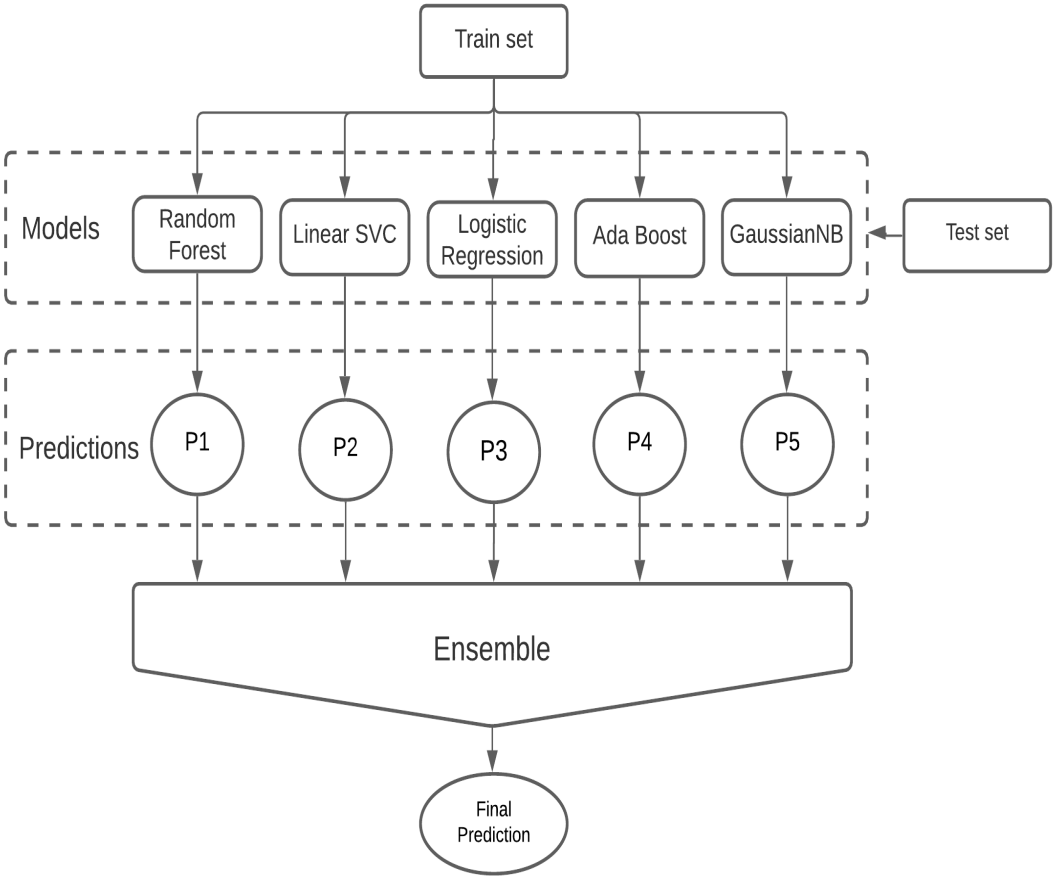

Figure 3.2: Ensemble Architecture used in non MRI data analysis

3.4 Design and Implementation

All the mentioned steps of the prototype are implemented using Python [20]. The algorithms and Ensemble architecture are implemented using the machine learning library Scikit-learn [21]. For additional calculation, implementation, and dataset analysis, Numpy [22] and Panda [23] are used. Finally, the architecture is implemented in the Google Colaboratory for reducing computational complexity.

3.5 Summary

This section explains the architecture of the proposed method. The overall architecture uses several machine learning models and the combined Ensemble Architecture.

Chapter 4

Implementation, Testing, and Result Analysis

4.1 Introduction

This section explains the architecture of the AD classification method, Dataset used and the evaluation process used in this research. The overall architecture uses the Ensemble Architecture approach.

4.2 Dataset

For this study, we used the OASIS [24] longitudinal dataset. OASIS is the Open Access Series of Imaging Studies (OASIS). This project aims to make neuroimaging data sets of the brain freely available to the scientific community. OASIS Dataset provides several independent variables including sociodemographic and clinical variables. Sociodemographic variables used in this experiment are:

- Gender: Male = 150, Female = 204.
- Age: Min = 60, Max = 98, Mean = 77.03.

- Years of Education (EDUC): Min = 6, Max =23, Mean = 14.7.
- Socioeconomic Status (SES): Upper = 7, Upper Middle = 74, Middle = 82, Lower Middle = 103, Lower = 88.

Clinical Variables are:

- Mini-Mental State Examination score (MMSE): The MMSE is a 30-point questionnaire widely used to assess cognitive impairment in clinical and research environments [25]. However, MMSE alone cannot be used for the classification of a disease.
- Clinical Dementia Rating (CDR): The CDR is a 5-point scale used to characterizing and tracking a patient's dementia level. CDR is estimated based on a semistructured interview of the subject and the caregiver (informant) and the clinician's clinical judgment.
- Estimated total intracranial volume (eTIV): The eTIV variable is the estimation of total intracranial brain volume.
- Normalize Whole Brain Volume (nWVB): The nWVB variable measures the volume of the whole brain.
- Atlas Scaling Factor (ASF): Atlas Scaling Factor is a computed scaling factor that transforms native-space brain and skull to the atlas target (i.e., the determinant of the transform matrix). [26]

In our test, we used Gender, Age, Years of education, Socioeconomic Status, Mini-Mental State Examination score, and Clinical Dementia Rating for the non-MRI test. In our Test with MRI data, we used all of these variables. We used the same data tuples for both tests.

4.3 Evaluation

We compared the result of the two experiments in different scales like Accuracy, F1 score, Recall, Precision. It is essential to measure the models in different areas to evaluate them from different perspectives. Accuracy is the percentage of correct predictions for the test data. The equations of Accuracy is,

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

Here, TP stands for True Positive, TN denotes True Negative, FP denotes False Positive, and FN denotes False Negative. However, Accuracy is not always the best scale to measure a model's performance. Precision tells us how many of the predicted positive results are actually positive.

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

Recall tells us how many of the positives are captured by our model. Recall is very important for a disease prediction model. Because we want to accurately identify as many positive patients as possible.

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

F1 score is used when we need to balance between Precision and Recall. F1 score tells us which model has the optimal false positive and true positive rates. The equation for measuring the F1 score is,

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

We analyzed the Area Under the Receiver Operating Characteristics to determine which model best distinguishes classes. The model with the highest Receiver Operating Characteristic (ROC) – Area Under the Curve (AUC) value has the best separability.

4.4 Results

We run our test for both MRI and non-MRI data. After that, we used different matrices to compare them.

4.4.1 MRI data results

We got the best Accuracy and a recall score from AdaBoost Classifier. It had an Accuracy score of 96.07%, a Recall of 0.96, a Precision of 0.96, and an F1 score of 0.96. The Accuracy score of Logistic Regression was 94.37%, a Recall score of 0.94, a Precision score of 0.94, and an F1 score of 0.94.

We got the best AUC score from Logistic Regression, 0.97, AdaBoost Classifier had an AUC score of 0.95. The Area Under the Curve (AUC) equals a classifier's probability of ranking a randomly chosen positive instance higher than a randomly chosen negative example. So the logistic Regression model has the best separability between positive and negative results.

The results of our hybrid model were not as good as the individual models. Our hybrid model scored an Accuracy score of 92.55%, a Recall score of 0.93, a Precision score of 0.93, an F1 score of 0.93, and an AUC score of 0.95.

Models	Accuracy	Precision	Recall	F1 score	AUC score
Random Forest	93.14%	0.93	0.93	0.93	0.91
Linear SVC	90.14%	0.90	0.90	0.90	0.92
Logistic Regression	94.37%	0.94	0.94	0.94	0.97
Ada Boost	96.07%	0.96	0.96	0.96	0.95
GaussianNB	86.00%	0.86	0.86	0.86	0.90
Ensemble Model	92.55%	0.92	0.92	0.92	0.95

Table 4.1: Accuracy, Precision, Recall, F1 score, and Area Under the Curve score of our models with MRI data.

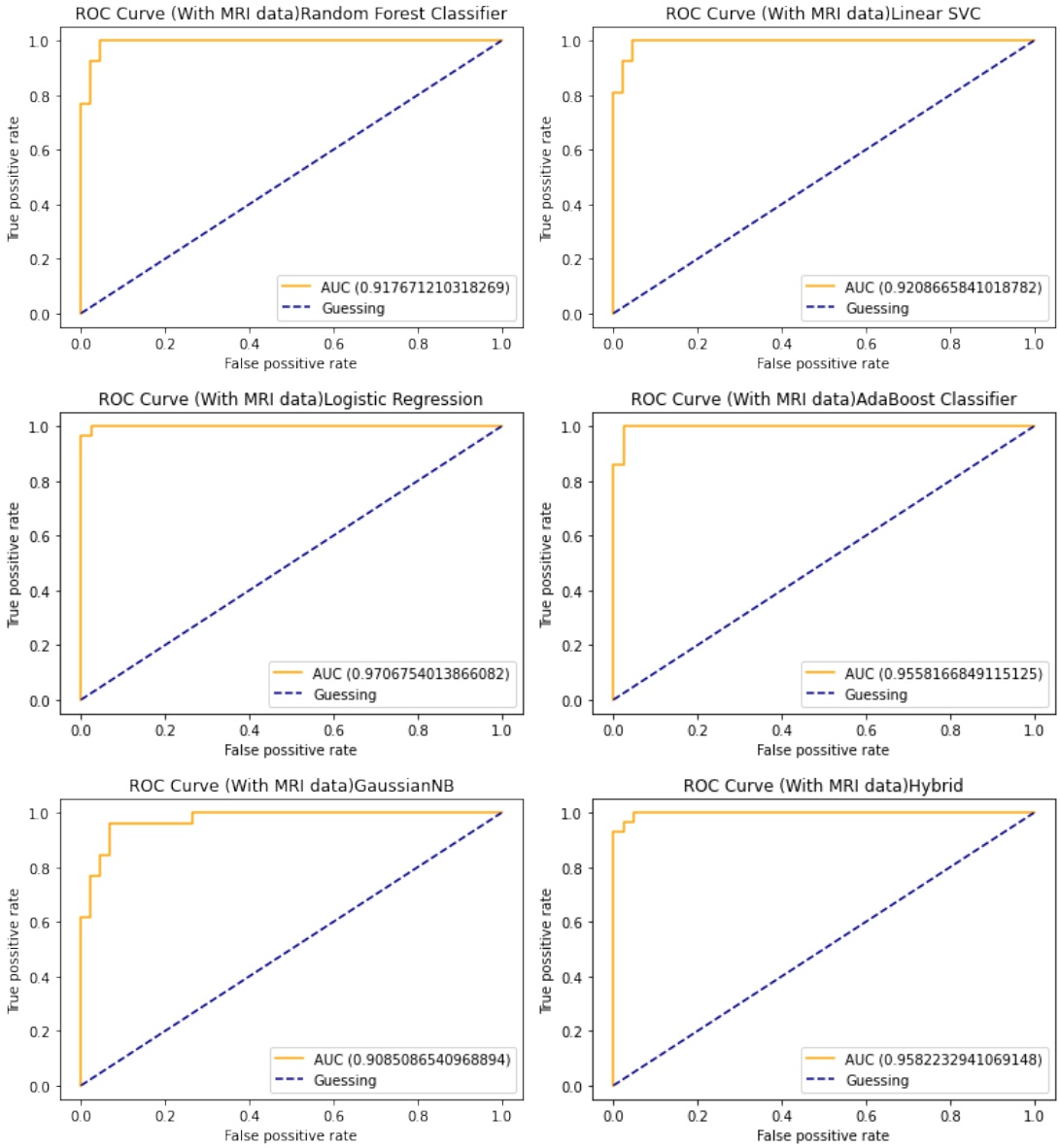

Figure 4.1: MRI data results, The Receiver Operating Characteristic curve (ROC) and with Area Under the Curve (AUC) score of MRI data for Random Forest, Linear SVC, Logistic Regression, Ada Boost, GaussianNB, Ensemble Model

4.4.2 Non-MRI data results

Our best performing model was the Random forest classifier when we trained our models without MRI data. This model scored an Accuracy score of 90.14%, a Recall of 0.90, a Precision of 0.90, and an F1 score of 0.90. The second-best performing models were LinearSVC and the AdaBoost classifier. Both models

achieved an Accuracy score of 89.14%, a Recall of 0.89, a Precision of 0.89, and an F1 score of 0.89. The AUC scores for these Random forest Classifiers, AdaBoost classifier and LinearSVC were 0.945 and 0.923 and 0.92, respectively.

In our training with MRI data KNeighbors Classifier performed very poorly. However, in this case, when trained without MRI data, this model performed better than before. It is a perfect example that modifying a dataset can improve the performance of a model.

We got our best result when we combined all the models. Our Hybrid model scored an Accuracy score of 93.37%, a Recall of 0.93, a Precision of 0.93, and an F1 score of 0.93. The AUC score for this hybrid model was 0.917.

Models	Accuracy	Precision	Recall	F1 score	AUC score
Random Forest	90.14%	0.90	0.90	0.90	0.94
Linear SVC	89.14%	0.89	0.89	0.89	0.92
Logistic Regression	88.73%	0.89	0.89	0.89	0.94
Ada Boost	89.14%	0.89	0.89	0.89	0.94
GaussianNB	88.73%	0.89	0.89	0.89	0.93
KNeighbors	83.10%	0.83	0.83	0.83	0.90
Ensemble Model	93.37%	0.93	0.93	0.93	0.91

Table 4.2: Accuracy, Precision, Recall, F1 score, and Area Under the Curve score of our models with non-MRI data.

Comparing this result with our previous results shows that GaussianNB and Hybrid models also performed better when trained without MRI data, like KNeighbors Classifier. Linear SVC performed the same as before. However, our overall Accuracy decreased. Previously our best model had 96.07% Accuracy, but when we removed all the MRI data, our Accuracy was 93.37%. So, we got an Accuracy drop of 2.7%. When it comes to detecting a patient, we always want to avoid False Negatives. So, we can consider the recall score for selecting our optimal model. The Recall decreased by 0.03, from 0.96 to 0.93, which is not a significant drop. The drop in the AUC score was 0.03. It dropped from 0.97 to 0.94.

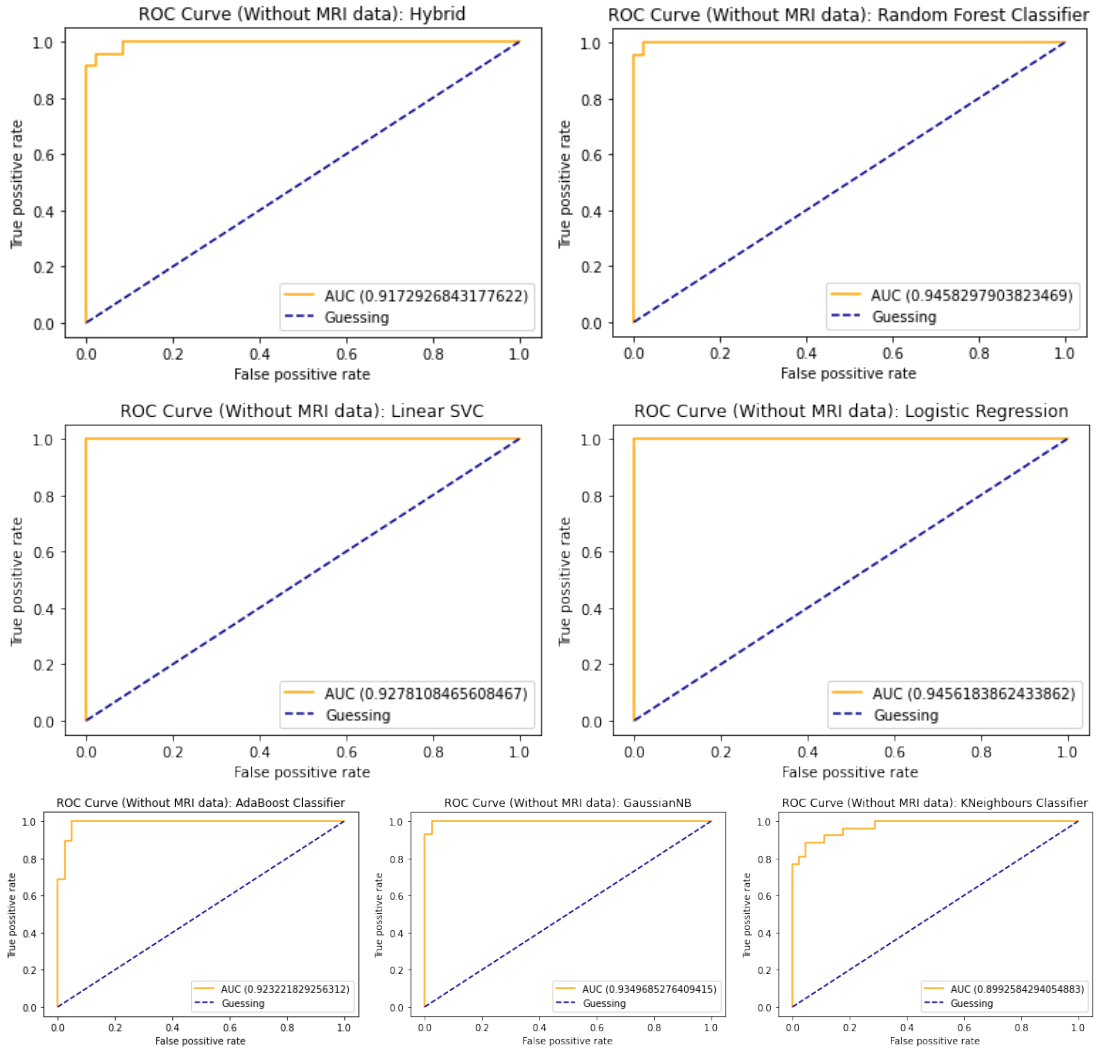

Figure 4.2: Non-MRI data results, The Receiver Operating Characteristic curve (ROC) and with Area Under the Curve (AUC) score of non-MRI data for our Ensemble Model, Random Forest, Linear SVC, Logistic Regression, Ada Boost, GaussianNB and KNeighbours classifier Model

4.5 Results analysis

From our result, we can see no considerable gap in efficiency between our two experiments, only about a 3% difference in Accuracy. MRI tests have significant importance for monitoring the disease and treatment. Some papers have found better Accuracy with MRI data than ours. Fulton et al. [7] found a 99.34% correct classification rate using MRI images and 91.3% accuracy in predicting CDR using non-MRI data. So, MRI classification is nearly perfect. So non-MRI

methods have about a 7% gap, which is significant. However, the main objective of our research was not to find an optimal way of predicting Alzheimer's disease but to show that the non- MRI methods are not too inefficient comparing to MRI data analysis methods.

Though the Non-MRI method has good results, it is not perfect. Some Demented patients may not be detected with this approach. It is recommended to do an MRI test if possible for better detection and avoid the risk of false-negative prediction. False-negative predictions can be dangerous sometimes when it comes to detecting a disease. However, the non-MRI methods can provide reliable results when there is a lack of diagnostic assessment or cannot afford an MRI test. This will reduce their initial cost by a margin. The non-MRI method can raise awareness in developing nations, as most of them are unaware of Alzheimer's disease.

These results can be further improved by counting more AD-related factors, like depression, head trauma, epilepsy, diabetes, stroke, and similar factors. We firmly believe our research will pave the way to raise awareness and cheap diagnosis of AD with excellent efficiency.

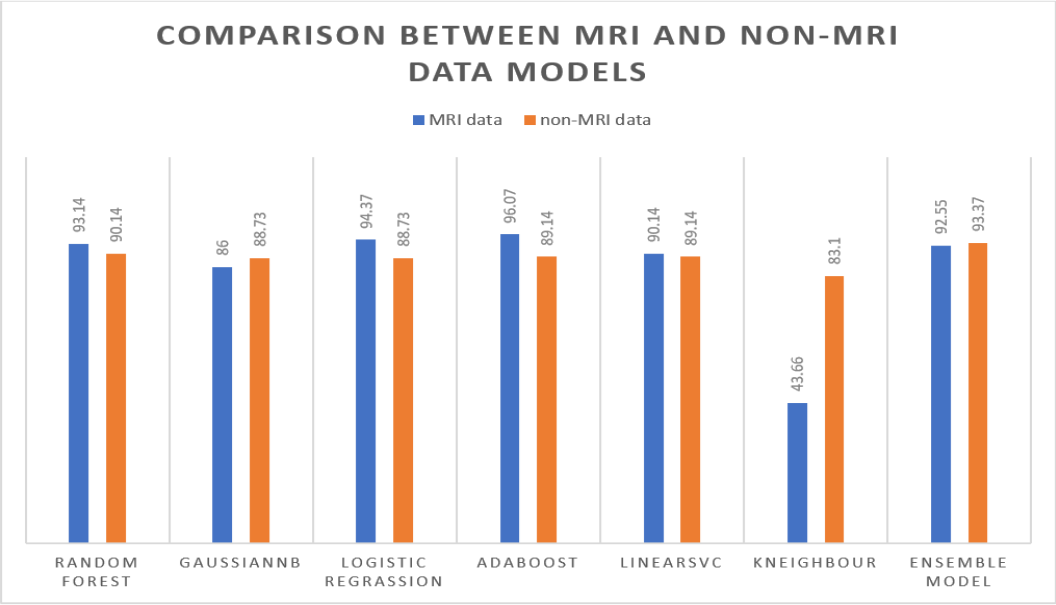

Figure 4.3: Bar-chart showing the performance of different model in both MRI and non-MRI dataset

4.6 Summery

From the evaluation analysis, It is proved that the non-MRI data is effective in predicting AD. Which is cheaper than MRI tests and it performs really well.

Chapter 5

Standards, Impacts and Milestones

5.1 Introduction

This section demonstrates the Standards, Impacts, Ethics, and Challenges of the thesis work. Then, the Constraints and Alternatives are illustrated. Finally, the Schedules, Tasks, and Milestones of the proposed work are presented.

5.2 Standards

We ensure that our thesis work will be sustainable for many years. Predicting AD is a very popular field of study and it will remain popular as the cure to AD is yet to be found [27]. Predicting AD at cheap cost is very important for the society and saves lives. It is also important for raising awareness among unaware people. We implemented the our methods with cutting edge machine learning technology. As our used resources will be available for more extended periods of time, we can say this thesis work will be sustainable.

5.3 Impacts on society

Our method can have a profound impact on society. Many people from low-income families suffer from AD without any treatment or care. However, a low-cost system will open doors for those people. Though cure is not available, the proper care can reduce the progression of AD and potentially save lives. As the AD is alarmingly increasing in developing countries [1], this method will be beneficial for the poor people living in those countries.

It can also be used as a tool for raising awareness. People believe that losing memory is a typical aging problem in our society. For this reason, many people from wealthy families also suffer from AD without any proper treatments. So it is a perfect tool for raising awareness as it is cheap and does not need any diagnosis instruments. Once these families are aware about the disease, they will be able to provide proper care and treatment.

5.4 Challenges

Although the field of AD research has come a long way, there are still several challenges in this field. The biggest challenge, of course, is to develop a cure for the disease. A lack of trained professionals is also a big challenge. Low awareness and misconception about the disease are also challenges moving forward, as people are reluctant to diagnose for AD.

5.5 Constraints

Different constraints such as design constraints, component constraints, and budget constraints are presented in this section. The overall structure is proposed based on training several machine learning models. We developed our system on Google Colaboratory to reduce hardware constraints. However, we still faced some issues with the free version of the Google Colaboratory. We could not run more than one process simultaneously, so it took more time to

train the models. A stable internet connection was also an issue, as models needed to train from the beginning if we were disconnected from the Google Colaboratory.

5.6 Timeline

Our thesis work timeline is divided into three segments based on the three semesters we got to complete our work. Each semester is four months and we got one year to complete our thesis. We have carried our work following the guidelines of our supervisor. In The first semester, we submitted a proposal and reviewed the related work of the thesis work. In the second semester, we implemented the model partially. Finally, we implemented the overall architecture in the third semester, benchmarked our results, and reported the overall work. In the meantime, we also wrote a conference paper which was accepted. Figure 7.1 contains the Gantt chart describing the work execution process of the thesis work.

5.7 Summary

This chapter briefly explains the standards, impacts, challenges constraints, schedules, tasks and milestones of the proposed work are demonstrated.

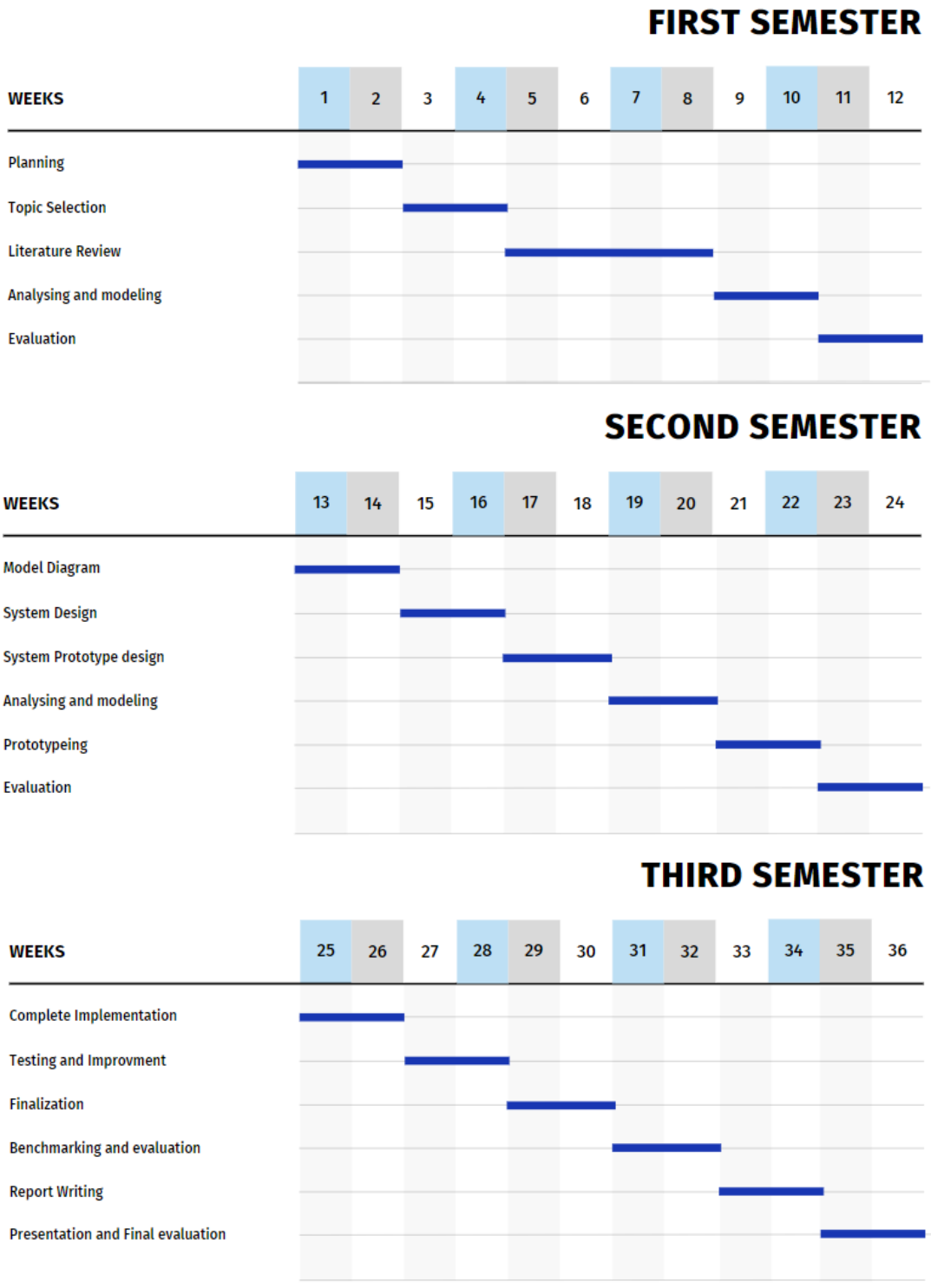

Figure 5.1: Gantt chart of the work execution process.

Chapter 6

Conclusion

6.1 Introduction

This paper experiments and evaluates a low-cost diagnostic model of AD. We tested and benchmarked both MRI and non-MRI data to measure the efficiency gap between these two methods. We also developed an Ensemble model that performs well with non-MRI data. This is the first research to perform this benchmark of this kind.

6.2 Future Works and Limitations

Our thesis is the first research work on low-cost AD prediction. This can be an extensive field of research. We have worked with a comparatively small dataset. This method can be tested for large datasets for better results. We strongly believe that "Predicting Alzheimer's Disease at Low Cost Using Machine Learning" is a research work that will pave the way for significant research on the cheaper diagnosis of AD

Bibliography

[1] C. P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, *et al.*, "Global prevalence of dementia: a delphi consensus study," *The lancet*, vol. 366, no. 9503, pp. 2112–2117, 2005.

[2] I. Raicher, M. M. Shimizu, D. Y. Takahashi, R. Nitrini, and P. Caramelli, "Alzheimer's disease diagnosis disclosure in brazil: A survey of specialized physicians' current practice and attitudes," *International Psychogeriatrics*, vol. 20, no. 03, 2007.

[3] K. Fiore, "Copyright issues hinder mmse use." `https://www.medpagetoday.com/neurology/dementia/52040?vpass=1`, Jun 2015.

[4] I. Raicher, M. M. Shimizu, D. Y. Takahashi, R. Nitrini, and P. Caramelli, "Alzheimer's disease diagnosis disclosure in brazil: a survey of specialized physicians' current practice and attitudes," *International Psychogeriatrics*, vol. 20, no. 3, pp. 471–481, 2008.

[5] A. R. Borenstein, C. I. Copenhaver, and J. A. Mortimer, "Early-life risk factors for alzheimer disease," *Alzheimer Disease & Associated Disorders*, vol. 20, no. 1, pp. 63–72, 2006.

[6] M. A. Lopes, S. R. Hototian, S. E. Bustamante, D. Azevedo, M. Tatsch, M. C. Bazzarella, J. Litvoc, and C. M. Bottino, "Prevalence of cognitive and functional impairment in a community sample in ribeirão preto, brazil," *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, vol. 22, no. 8, pp. 770–776, 2007.

[7] L. V. Fulton, D. Dolezel, J. Harrop, Y. Yan, and C. P. Fulton, "Classification of alzheimer's disease with and without imagery using gradient boosted machines and resnet-50," *Brain sciences*, vol. 9, no. 9, p. 212, 2019.

[8] G. Battineni, N. Chintalapudi, F. Amenta, and E. Traini, "A comprehensive machine-learning model applied to magnetic resonance imaging (mri) to predict alzheimer's disease (ad) in older subjects," *Journal of Clinical Medicine*, vol. 9, no. 7, p. 2146, 2020.

[9] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of alzheimer's disease with deep learning," in *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*, pp. 1015–1018, IEEE, 2014.

[10] Y. Lin, K. Huang, H. Xu, Z. Qiao, S. Cai, Y. Wang, L. Huang, A. D. N. Initiative, *et al.*, "Predicting the progression of mild cognitive impairment to alzheimer's disease by longitudinal magnetic resonance imaging-based dictionary learning," *Clinical Neurophysiology*, vol. 131, no. 10, pp. 2429–2439, 2020.

[11] C. Cochrane, D. Castineira, N. Shiban, and P. Protopapas, "Application of machine learning to predict the risk of alzheimer's disease: An accurate and practical solution for early diagnostics," *arXiv preprint arXiv:2006.08702*, 2020.

[12] X. Hong, R. Lin, C. Yang, N. Zeng, C. Cai, J. Gou, and J. Yang, "Predicting alzheimer's disease using lstm," *IEEE Access*, vol. 7, pp. 80893–80901, 2019.

[13] G. Lee, K. Nho, B. Kang, K.-A. Sohn, and D. Kim, "Predicting alzheimer's disease progression using multi-modal deep learning approach," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[14] M. Subramoniam *et al.*, "Deep learning based prediction of alzheimer's disease from magnetic resonance images," *arXiv preprint arXiv:2101.04961*, 2021.

[15] S. F. Eskildsen, P. Coupé, D. García-Lorenzo, V. Fonov, J. C. Pruessner, D. L. Collins, A. D. N. Initiative, *et al.*, "Prediction of alzheimer's disease in

subjects with mild cognitive impairment from the adni cohort using patterns of cortical thinning," *Neuroimage*, vol. 65, pp. 511–521, 2013.

[16] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, V. Calhoun, A. D. N. Initiative, *et al.*, "Deep residual learning for neuroimaging: An application to predict progression to alzheimer's disease," *Journal of neuroscience methods*, vol. 339, p. 108701, 2020.

[17] M. Nguyen, T. He, L. An, D. C. Alexander, J. Feng, B. T. Yeo, A. D. N. Initiative, *et al.*, "Predicting alzheimer's disease progression using deep recurrent neural networks," *NeuroImage*, vol. 222, p. 117203, 2020.

[18] A. Payan and G. Montana, "Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks," *arXiv preprint arXiv:1502.02506*, 2015.

[19] B. Lei, M. Yang, P. Yang, F. Zhou, W. Hou, W. Zou, X. Li, T. Wang, X. Xiao, and S. Wang, "Deep and joint learning of longitudinal data for alzheimer's disease prediction," *Pattern Recognition*, vol. 102, p. 107247, 2020.

[20] G. Van Rossum *et al.*, "Python," 1991.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[22] S. Van Der Walt, S. C. Colbert, and G. Varoquaux, "The numpy array: a structure for efficient numerical computation," *Computing in science & engineering*, vol. 13, no. 2, pp. 22–30, 2011.

[23] W. McKinney *et al.*, "pandas: a foundational python library for data analysis and statistics," *Python for high performance and scientific computing*, vol. 14, no. 9, pp. 1–9, 2011.

[24] "Oasis brains - open access series of imaging studie." `https://www.oasis-brains.org/#data`. Accessed: 11-May-2021.

[25] V. C. Pangman, J. Sloan, and L. Guse, "An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice," *Applied Nursing Research*, vol. 13, no. 4, pp. 209–213, 2000.

[26] R. L. Buckner, D. Head, J. Parker, A. F. Fotenos, D. Marcus, J. C. Morris, and A. Z. Snyder, "A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume," *Neuroimage*, vol. 23, no. 2, pp. 724–738, 2004.

[27] B. Heigle, A. Khan, R. Ottwell, and M. Vassar, "Use of exaggerated language in news stories to describe drugs for treatment of alzheimer's disease," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 6, no. 1, 2020.