



**North South University**  
Dept. of Electrical and Computer Engineering

**Title:** A Comparison of Machine Learning Algorithms for Predicting Alzheimer's Disease

**Group member name & ID:**

<b>Name:</b> Al Sabri Bhuiyan	<b>ID:</b> 1812098042
<b>Name:</b> Farhan Shahriar	<b>ID:</b> 1812575042
<b>Name:</b> Nafisa Tasnim Rifha	<b>ID:</b> 1831084642

**Supervisor:** Dr. Mohammad Monirujjaman Khan  
Associate Professor  
Department of Electrical and Computer Engineering  
North South University

**Date of Submission:** 29th November, 2021

## **1. Background study of the research topic:**

The most frequent kind of dementia is known as Alzheimer's disease. It is a gradual condition that begins with modest memory loss and may proceed to loss of capacity to converse and respond to the surroundings. Alzheimer's disease affects brain regions that govern cognition, memory, and language. It can have a significant impact on a person's capacity to carry out everyday duties [1]. Scientists do not yet completely comprehend what causes Alzheimer's disease in the majority of people. The causes are most likely a mix of age-related brain changes, as well as genetic, environmental, and lifestyle variables. The significance of each of these factors in increasing or lowering the risk of Alzheimer's disease varies from person to person. Alzheimer's disease is a brain disorder that develops with time. It is characterized by abnormalities in the brain, such as amyloid plaques and neurofibrillary, or tau, tangles, which cause neuronal and neurofibrillary connections to be lost. These and other changes have an impact on a person's capacity to recall and think, as well as their ability to live independently. Alzheimer's disease is classified into two types: early-onset and late-onset. There is a hereditary component to both sorts [2].

Alzheimer's disease was described for the first time in 1906. Auguste D, a woman with substantial memory loss, unjustified suspicions about her family, and other increasing psychological problems, was treated by German physician Alois Alzheimer, a pioneer in correlating symptoms to microscopic brain alterations. At autopsy, he discovered severe shrinkage and aberrant deposits in and around nerve cells in her brain.

In 1910, Emil Kraepelin, a German psychiatrist who worked with Dr. Alzheimer, coined the term "Alzheimer's Disease" in the ninth edition of his book *Psychiatrie* [3].

Around 55 million individuals worldwide suffer from dementia, with over 60% of them residing in low- and middle-income nations. This figure is predicted to climb to 78 million in 2030 and 139 million in 2050, as the share of older people in the population grows in practically every country. Every year, about 10 million new cases are reported.

There is no cure for dementia at this time. Anti-dementia drugs and disease-modifying therapies currently on the market have limited effectiveness and are mostly designated for Alzheimer's disease, however a plenty of novel treatments are in different stages of clinical testing. The primary objective of dementia care, on the other hand, is early diagnosis in order to facilitate early and effective management [4].

## **2. Description of the problem being solved (Problem statement):**

The majority of Bangladeshis are ignorant of the Alzheimer's disease. As a result of our efforts, we will be able to assist our country's residents in understanding what it is and how it affects them. Our doctors will be able to treat patients more successfully if Alzheimer's disease can be recognized in a short amount of time and at an early stage in their bodies. As a result, the number of people diagnosed with Alzheimer's disease is expected to decrease steadily. As a consequence of our study, we hope to be able to detect Alzheimer's disease at an early stage.

## **3. Review of existing similar systems:**

Many scholars have advocated a comprehensive study on the categorization and diagnosis of Alzheimer's disease. A quick summary of the relevant work is included in this section.

The structural MRI was utilized by Klöppel et al. (2008) to identify Alzheimer's disease from healthy controls at an early stage. SVM was used to MRI by the authors in order to reliably identify illness while separating it from normal aging. This study used pathologically proved data

sets as an input for successful categorization, which were obtained from several centres. Finally, the suggested technique was tested using normalized datasets from 67 Alzheimer's patients and 91 healthy controls from various scans. Using whole brain imaging and leave one out cross validation, 96 percent of pathologically validated alzheimer's disease patients were properly categorized. By merging data from many sites, the suggested study demonstrated generality; nevertheless, the data set is too small to make meaningful comparisons with other methodologies.

Although the use of many single biomarkers yields promising findings, they are intended to define group differences rather than to classify individuals. D. Zhang et al. (2011) developed a strategy for distinguishing between healthy and Alzheimer's disease participants by combining all three biomarkers for Alzheimer's disease diagnosis, such as MRI (magnetic resonance imaging), PET (positron emission tomography), and CSF (Cerebrospinal fluid).

The researchers used a baseline data set of 202 cases, 51 of which were Alzheimer's disease, 99 MCI (mild cognitive impairment), and 52 healthy controls. MRI, PET, and CSF tests, as well as their combination, were completed utilizing 10-fold cross validation.

The combination of these modalities provided a classification accuracy of 93.2 percent, with 93 percent sensitivity and 93.3 percent specificity, while separate tests yielded the maximum accuracy of 86.5 percent. For whatever number of brain areas picked, the authors stated that the multimodal classification technique (using all MRI, PET, and CSF) shows constant improvement and is more robust than those employing individual modality. CSF and PET offer the most complimentary information, whereas MRI and PET have the most comparable information for categorization, according to these findings.

Furthermore, it should be highlighted that the amount of data available for each individual patient across all modalities is insufficient to allow for meaningful categorization. This study makes no mention of missing values knowledge or how they are handled. Another significant flaw in this article is the imbalance of social classes [5].

#### **4. Project Proposal:**

To overcome the challenges listed above, we aim to use machine learning to provide a solution. Because machine learning has proven to be quite effective in the medical industry. Machine learning is being used in a variety of healthcare settings, from case management of common chronic conditions to utilizing patient health data in conjunction with environmental variables such as pollution exposure and weather. ML technology can assist healthcare practitioners in developing accurate medication treatments suited to individual features by crunching enormous amounts of data. It may also be used to show and educate patients about illness processes and consequences as a result of various treatment options. It has the potential to improve hospital and health-care system efficiency while lowering costs [6]. ML is an expert at diagnosing problems; in fact, this is one of her strongest skills. Many forms of cancer and genetic illnesses are difficult to diagnose, but ML could manage many of them in their early stages [7].

In medical diagnostics, machine learning models are becoming routinely employed. Here, we propose to compare different machine learning performances to diagnose Alzheimer's syndrome. According to studies, people can discover Alzheimer's disease early and begin treatment. They must precisely forecast the progression of the disease from moderate condition to Dementia, in order to do so.

Machine learning technology can aid in the precise prediction of Alzheimer's disease in its early stages. (There are numerous machine learning systems, but their predictions are erratic and wrong.) They also have concerns with overfitting and underfitting.

As a result, we built a model that can detect Alzheimer's disease early using machine learning to assist medical technicians. It will determine whether or not someone has Alzheimer's disease [8].

### **5. Objective of the project:**

- Our main objective is to diagnose Alzheimer's syndrome using machine learning.
- We will implement three different machine learning algorithms in this project. Which are: Ada boost classifier, Extra Trees classifier and Support vector machine classifier.
- Following that, we will do a comparison analysis to determine which algorithm provides the best outcomes.

### **6. Output of the project or expected results of the project:**

The project's main output is to create a journal based on the findings of our analysis. Where We will also discuss our techniques during the full analysis. Finally, we'll strive to have our journal published.

And the project's expected result is to provide the best level of accuracy compared to any current equivalent system.

### **7. Feasibility study:**

Machine learning (ML) is the study of computer systems that learn via inference and patterns without being explicitly programmed using algorithms and statistical models. Over time,

machine learning algorithms learn and develop on their own. It discovers approaches, trains models, and utilizes the learnt methodology to automatically decide the output. Machine learning systems may also adapt to changes in their surroundings [8].

A model is a machine learning system that has been taught to recognize particular sorts of patterns with the help of an algorithm. That is, it analyses data and uncovers latent structures in a dataset. The formula that depends on the input and output functions and applies it to fresh data to anticipate the response is determined by the feature extraction and known replies of a dataset. As a result, the model's algorithm trains on a set of data, creates a method for predicting the outcome, and saves the technique for future use [8].

For two-group classification issues, a support vector machine (SVM) is a supervised machine learning model that employs classification techniques. Support vector machine (SVM) is a rapid and trustworthy classification technique that works well with little amounts of data. SVMs are a collection of supervised learning algorithms used to solve classification and regression issues [8].

A line that divides the input variable space is called a hyperplane. A hyperplane is chosen in SVM to optimally segregate the points in the input variable space by their class, which can be either class 0 or class 1. In two dimensions, this may be seen as a line, and let's assume that this line can entirely separate all of our input points. The SVM learning method discovers the coefficients that produce the optimal hyperplane separation of the classes. The margin is defined as the distance between the hyperplane and the nearest data points. The line with the widest margin is the best or optimal hyperplane for separating the two classes. Only these points matter when it comes to defining the hyperplane and building the classifier. The support vectors are the points that make up the support vectors. They help to define or sustain the hyperplane. In

practice, to discover the values for the coefficients that maximize the margin, an optimization technique is utilized [9].

Adaptive Boosting is abbreviated as Adaboost. Because each model is developed individually, bagging is a parallel ensemble. Boosting, on the other hand, is a sequential ensemble in which each model is constructed by correcting the prior model's misclassifications.

Bagging mostly simple voting, in which each classifier votes for a final outcome— one that is determined by the majority of the parallel models; boosting entails 'weighted voting,' in which each classifier votes for a final outcome that is determined by the majority— but the sequential models were built by assigning higher weights to misclassified instances of the previous models [10].

Short decision trees are utilized with Adaboost. Following the creation of the first tree, the tree's performance on each training instance is used to weight how much attention the following tree should give to each training instance. Training data that is difficult to predict is given more weight, whereas examples that are easy to anticipate are given less weight. Models are built in a sequential order, with each tree updating the weights on the training examples that impact the learning of the following tree in the sequence. After all of the trees have been formed, new data predictions are produced, and each tree's performance is weighted based on how accurate it was on training data. Because the algorithm focuses so much on fixing errors, it's critical to have clean data that's free of outliers [9].



## **8. Solutions adopted and the reasons for that:**

One of the primary reasons for adapting machine learning as a solution is that it can identify and detect diseases and symptoms that are normally difficult to diagnose. This can include anything from malignancies that are difficult to detect in their early stages to other hereditary illnesses. Even one of the above great instances of how merging cognitive computing with the genome-based tumour sequence might aid in generating a quick diagnosis is IBM Watson Genomics [11].

The SVM's key benefit is that it can easily overcome the high dimensionality problem, which occurs when there are a large number of input variables in comparison to the number of available data. Furthermore, because the SVM technique is data-driven and achievable without a theoretical foundation, it may offer high discriminative power for classification, especially with small sample sizes. This technology has recently been employed in clinical settings to improve illness detection methods. Furthermore, SVM has shown to be effective in handling bioinformatics classification challenges. In this work, we used SVM to predict patient's medication adherence using common factors that are relatively easy to gather [12].

According to recent studies, classification is the most extensively used ML problem in the medical field, and solutions based on the AdaBoost algorithm make up a substantial portion of the research. Alzheimer's illness, diabetes, hypertension, and different malignancies are among the clinical uses. Non-clinical examinations of self-reported mental health and subhealth status are also available. Chronic fatigue and weakness characterize the latter, which frequently leads to eventual illness. Unstructured, user-generated information from online health forums was utilized in these non-clinical ways. AdaBoost has also been utilized as a pre-

processing technique for automatically selecting the most significant features from large datasets [13].

#### **9. Working steps:**

- Week-1: Collecting the dataset
- Week-2: Reading the existing papers based on the related topic and dataset.
- Week-3: Start pre-processing the data.
- Week-4: Feature selection of the data.
- Week-5: Implementing algorithm-1.
- Week-6: Implementing algorithm-2.
- Week-7: Implementing algorithm-3.
- Week-8: Start writing report.

#### **10. Major Milestones:**

- Analysing the existing system.
- Dataset connection
- Pre-processing the data
- Feature selection
- Training
- Testing

## **11. Research methodology:**

11.1. Dataset: The system's major purpose is to forecast Dementia in different patients based on a variety of factors. The technique was developed using the longitudinal Magnetic Resonance Imaging (MRI) data from OASIS. Name of the dataset is “MRI and Alzheimers (Magnetic Resonance Imaging Comparisons of Demented and Nondemented Adults)”.

In the realm of machine learning, the OASIS dataset is rather tiny, with 374 rows and 15 columns. Here different characteristics is listed. The final conclusion took into account the person's gender (M/F), age, years of education (EDUC), socioeconomic status (SSE), mini-mental state examination (MMSE), estimated total intracranial volume (eTIV), normalized whole brain volume (nWBV), and Atlas scaling factor (ASF). The categorical aspects of the dataset that have been taken into account for the proposed model to predict Alzheimer's disease, as well as their descriptions, are: group, hand, and M/F. The majority of the data are numerical. All of these terms explain the patient's state and help the ML system determine the stage of dementia.

11.2. ML Algorithms: In this project we proposed to implement three different machine learning algorithms for a comparison analysis. All the proposed algorithms are described below:

i) Adaboost classifier: Boosting is a generic technique for increasing the accuracy of any learning system. Boosting is a strategy of combining rough and relatively incorrect rules of thumb to produce a very accurate prediction rule that has been proven to be useful. The AdaBoost algorithm has undergone extensive theoretical and practical research.

Freund and Schapire released the AdaBoost algorithm in 1995, which solved many of the preceding boosting techniques' practical issues. As a result, it is used in this framework [15].

A boost classifier is a classifier in the form of 
$$F_T(x) = \sum_{t=1}^T f_t(x)$$

with each  $f_t$  representing a weak learner, that receives an item  $x$  as input and outputs a value indicating the object's class. The sign of the weak learner output, for example, specifies the predicted object class in the two-class issue, whereas the absolute value indicates the confidence in that classification. Similarly, if the sample belongs to a positive class, the  $T^{\text{th}}$  classifier is positive; otherwise, it is negative.

For each sample in the training set, each weak learner generates an output hypothesis,  $h(x_i)$ . A weak learner is chosen and given a coefficient  $\alpha_t$  at each iteration  $t$  so that the total training error  $E_t$  of the resultant  $t$ -stage boost classifier is minimized.

$$E_t = \sum_i E[F_{t-1}(x_i) + \alpha_t h(x_i)]$$

Here,  $F_{t-1}(x)$  represents the boosted classifier that was developed up to the previous step of training,  $E(F)$  represents some error function, and  $f_t(x) = \alpha_t h(x)$  represents the weak learner that is being considered for inclusion in the final classifier.

At each iteration of the training process, a weight  $W_{i,t}$  equal to the current error  $E(F_{t-1}(x_i))$  on that sample is applied to each sample in the training set. These weights can be used to guide the weak learner's training; for example, decision trees that prefer dividing sets of samples with high weights can be generated [16].

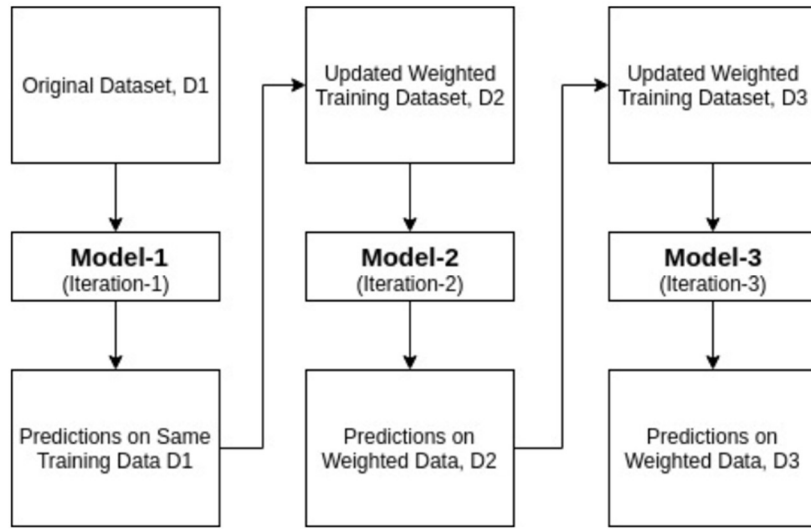


Fig: Adaboost Classifier block diagram

ii) Extra Trees Classifier: Extremely Randomized Trees Classifier (Extra Trees Classifier) is a form of ensemble learning technique that outputs a classification result by aggregating the outcomes of several de-correlated decision trees collected in a "forest." It is conceptually identical to a Random Forest Classifier, with the exception of how the decision trees in the forest are  $Entropy(S) = \sum_{i=1}^c -p_i \log_2(p_i)$  constructed. The Extra Trees Forest's Decision Trees are all made from the original training sample. Then, at every test node, each tree is given a random sample of k features from the feature set, from which it must choose the best feature to split the data according to certain mathematical criteria (typically the Gini Index). Multiple de-correlated decision trees are created from this random sample of features. The information collected is employed as a decision criterion in this case. We begin by calculating the data's entropy. The entropy is calculated using the formula

where  $c$  is the number of unique class labels and  $p_i$  is the proportion of rows having the output label  $i$  [17].

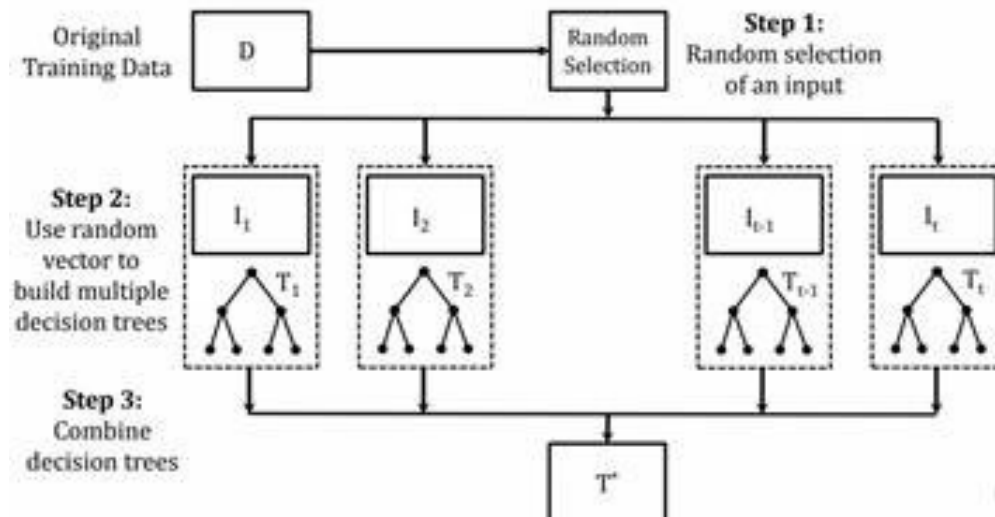


Fig: Extra Trees Classifier block diagram

iii) Support vector machine classifier: The Support Vector Machine (SVM)

is a Supervised Machine Learning Algorithm that is used for classification and regression. It is most commonly used for classification, although it can also be beneficial for regression. SVM basically finds a hyper-plane that separates the different sorts of data. This hyper-plane is nothing more than a line in two-dimensional space. Each data item in the dataset is plotted in an  $N$ -dimensional space in SVM, where  $N$  is the number of features/attributes in the data. Next, determine the best hyperplane for separating the data. As a result, you must have realized that SVM can only conduct binary classification by default (i.e., choose between two classes). For multi-class situations, however, there are a variety of approaches to employ.

For linearly separable data, SVM performs quite well without any adjustments. Any data that can be shown in a graph and divided into groups using a straight line is referred to be linearly separable data. For non-linearly separable data, we employ Kernelized SVM. Let's say we have some data in one dimension that is non-linearly separable. This data may be transformed into two dimensions, where it will be linearly separable in two dimensions. Each 1-D data point is mapped to a matching 2-D ordered pair in this way.

So we can simply move any non-linearly separable data in any dimension to a higher dimension and then make it linearly separable. This is a massive and all-encompassing metamorphosis.

A kernel is nothing more than a comparison of data points. In a kernelized SVM, the kernel function tells you how similar two data points in the original feature space are to points in the newly converted feature space.

There are many kernel functions to choose from, but two of the most common are:

Radial Basis Function (RBF): The distance between the vectors and the original input space is an exponentially declining function of the similarity between two points in the modified feature space, as illustrated below. In SVM, the default kernel is RBF.

$$K(x, x') = \exp(-\gamma ||x - x'||)$$

Polynomial Kernel: The degree parameter of the Polynomial kernel influences the model's complexity and computing cost of the transformation [18].

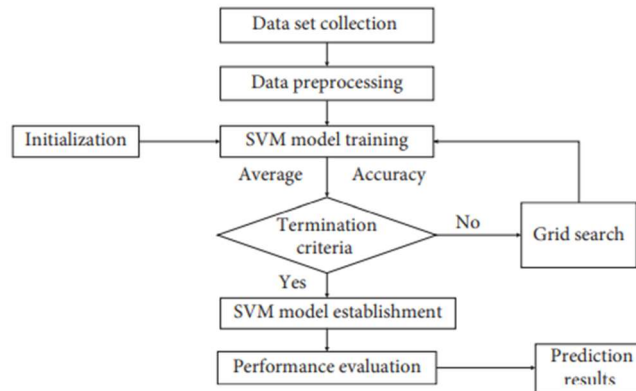


Fig: Support Vector Machine Classifier block diagram

11.3. Result Analysis: We'll use a confusion matrix to analyse our results. A confusion matrix, also known as an error matrix, is a special table structure that permits visualization of the performance of an algorithm, often a supervised learning one, in the field of machine learning and specifically the issue of statistical classification (in unsupervised learning it is usually called a matching matrix). The examples in an actual class are represented by each row of the matrix, whereas the instances in a predicted class are represented by each column, or vice versa - both variations are documented in the literature. The term comes from the fact that it's simple to tell whether the system is mixing up two types (i.e., commonly mislabelling one as another).

It's a unique type of contingency table, with two dimensions ("actual" and "predicted") and identical sets of "classes" in each (each combination of dimension and class is a variable in the contingency table).

There are four potential outcomes that might occur when comparing the actual classification set to the predicted classification set in each given column. One, a true positive result occurs when the actual classification is positive and the expected classification is positive (1,1), indicating that



the positive sample was accurately detected by the classifier. Two, a false negative result occurs when the actual classification is positive but the projected classification is negative (1,0), since the classifier wrongly identifies the positive sample as negative. Third, a false positive result occurs when the actual classification is negative and the projected classification is positive (0,1). This occurs because the classifier wrongly identifies the negative sample as positive. Fourth, a true negative result occurs when the actual classification is negative and the anticipated classification is negative (0,0), since the classifier successfully identifies the negative sample.

The four types of results outlined above (true positives, false negatives, false positives, and true negatives), as well as the positive and negative classifications, are used in the template for any binary confusion matrix. The four outcomes can be expressed as follows in a 2×2 confusion matrix:

		Predicted condition	
		Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

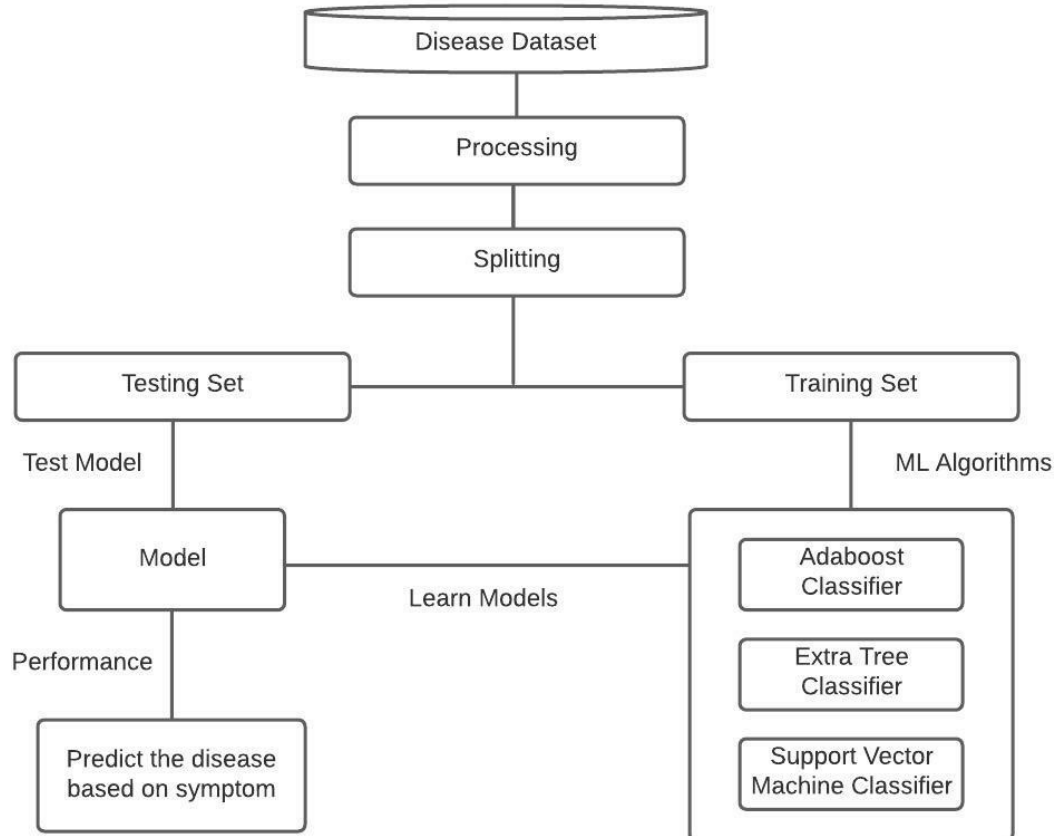
Now, Precision is the closeness of the measurements to each other, whereas accuracy is the proximity of the measurements to a given value in a set of measurements.

The precision formula is:  $PPV = \frac{TP}{TP+FP} = 1 - FDR$

Here FDR stands for False Discover Rate.

And the Accuracy formula is:  $ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN}$  [19]

## 12. Detailed diagrams for the complete system and all subsystems:



We'll start by gathering our preferred data set, which we can get at [kaggle.com](https://www.kaggle.com/). Following that, we will begin data pre-processing and analysis of the missing data. We'll then proceed to feature selection. Finally, we'll train and test the system using three distinct machine learning algorithms. Finally, we will implement it.

### **13. Explanation of the functioning of the complete system, and all subsystems:**

Data collection: Our dataset is on Alzheimer's disease. We collected our dataset from

[https://www.kaggle.com/jboysen/mri-and-alzheimers?select=oasis\\_longitudinal.csv](https://www.kaggle.com/jboysen/mri-and-alzheimers?select=oasis_longitudinal.csv)

Data Pre-processing: Data pre-processing is converting raw data into well-formed data sets in order to use data mining methods. Raw data is frequently incomplete and formatted inconsistently. The success of every project involving data analytics is directly proportional to the quality of data preparation. Data preparation is crucial in machine learning (ML) procedures to ensure that big datasets are prepared in such a manner that the data they contain can be processed and analysed by learning algorithms [20].

Feature selection: When creating a predictive model, feature selection is the process of minimizing the number of input variables.

The number of input variables should be reduced to lower the computational cost of modelling and, in some situations, to increase the model's performance.

The relationship between each input variable and the goal variable is evaluated using statistics, and the input variables having the strongest link with the target variable are selected. Although the choice of statistical measures depends on the data type of both the input and output variables, these approaches can be quick and successful [21].

Training: Data is used to train machine learning algorithms. They use the training data to form associations, gain insight, make judgments, and assess their confidence. The model works better when the training data is good. In reality, the quality and amount of your machine learning training data is just as important as the algorithms themselves in determining the success of your data project.

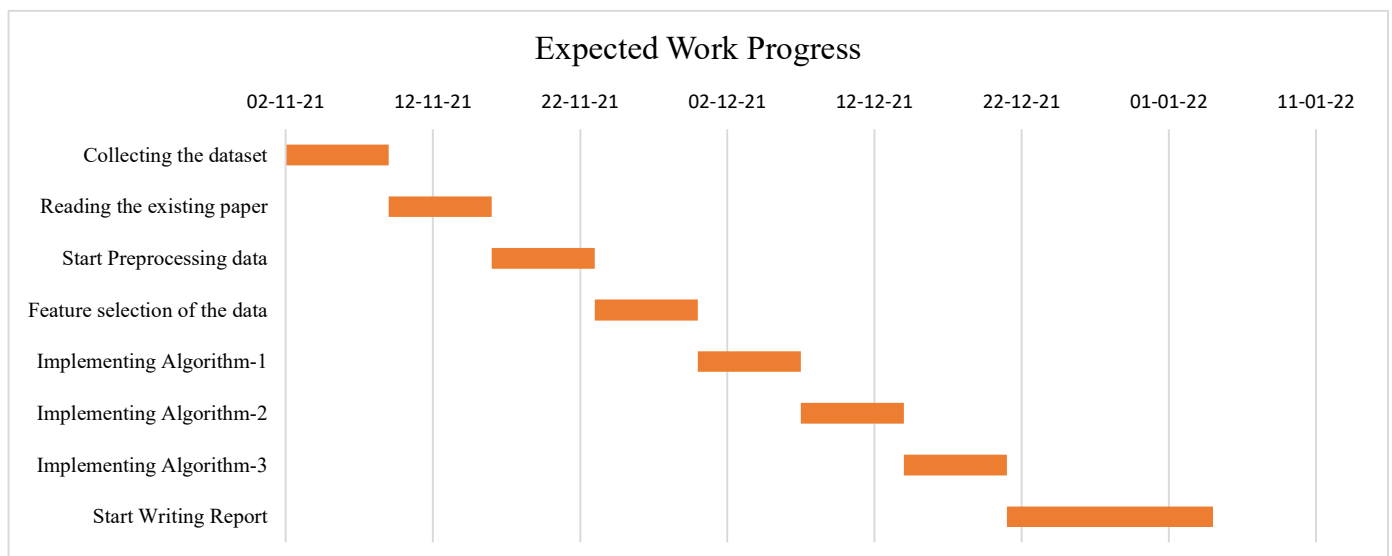
Testing: To create an ML algorithm, we'll need both training and testing data. A model is normally tested on a test set after it has been trained on a training set. These sets are frequently derived from the same larger dataset, albeit the training set should be tagged or enriched to improve the algorithm's confidence and accuracy [22].

Implementation: After completing all of the steps, we will proceed to the execution phase in order to obtain the desired outcome.

#### 14. Figure showing inputs and outputs:



#### 15. MS Project charts including Gantt Charts showing the expected timeline of progress or milestones:



## **16. Required software tools:**

Anaconda Navigator and Jupyter Notebook will be used to carry forward the model training and validation. And we will be using Python as our programming language.

Anaconda is a Python programming language distribution aimed for simplifying package management and deployment in scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, and so on). Data-science software for Windows, Linux, and macOS are included in the release.

Jupyter Notebook is an online tool that allows you to create and share documents with code, visuals, and text. It's useful for data science, statistical modeling, machine learning, and a variety of other tasks. We want to use this because Jupyter Notebook can connect to a variety of kernels, allowing you to program in a variety of languages. A Jupyter kernel is a program that handles a variety of requests (code execution, code completions, and inspection) and responds.

Python is a dynamically semantic, interpreted, object-oriented high-level programming language. Its high-level built-in data structures, together with dynamic typing and dynamic binding, making it perfect for Rapid Application Development and as a scripting or glue language for connecting existing components. Python's concise, easy-to-learn syntax prioritizes readability, which lowers software maintenance costs. Modules and packages are supported by Python, which fosters program modularity and code reuse.

## **17. Target Population:**

The people of Bangladesh are our target audience. People in Bangladesh are largely unaware of the illness. Through our efforts, we will be able to assist the citizens of our nation in

comprehending what it is and how it affects them. If Alzheimer's disease can be detected in a short period of time and at an early stage in a person's body, our doctors will be able to treat the patient more effectively. As a result, the number of people diagnosed with Alzheimer's disease will steadily decline. As a result of our research, we anticipate being able to identify Alzheimer's disease at an early stage.

#### **18. What makes the solution an ‘innovation’:**

Until now, every paper regarding our topic that we've read has had an accuracy rate of less than 90%. However, we are working with an extremely efficient dataset in our scenario. As a result, we'll aim to create a system that can help people. We'll use multiple models in this case. This will be a milestone for our project if we can get the highest accuracy while using multiple models. As a result, we might conclude that our solution is innovative.

#### **19. How the project will become sustainable:**

No adjustments will be required in the future for any of our models. Therefore, our project is sustainable. Since this is a software-based project, it will automatically update if the software changes in the future.

As a result, we do not need to be concerned about the project's long-term viability.

#### **20. Is there any possibility to scale up the project:**

Of course, the project may be scaled up. Right now, we're just getting started with the models for our project. We can build apps connected to it in the future and increase our work. We may also

use a larger dataset to broaden our scope of study. We can implement six algorithms and increase our work in the future because we are just implementing three algorithms today.

**21. Is there any opportunity of income generation from the project:**

Yes, there is an opportunity of income generation from this project.

To do so, we'll need to create a project-specific application. Then, in the future, if we prepare a patent, we will be able to earn money from it.

**22. How people will be benefit from the project:**

We shall aim to anticipate Alzheimer's disease at an early stage in this project. People have a higher chance of benefiting from therapy if they are diagnosed with Alzheimer's disease early. Individuals who receive an early diagnosis are eligible for a greater range of clinical studies, which enhance research while also potentially providing medical advantages. Receiving an early Alzheimer's diagnosis might help alleviate concerns regarding the cause of people's symptoms. They and their families may also make the most of their time together by making use of services and support programs.

**23. What are the risks? How the risks will be managed:**

We have no risks except that after implementing the algorithms, we may yield less accuracy.

If this occurs, we will manipulate the data and implement even more algorithms. As a result, the risk will be managed.

#### **24. Disabled will benefit:**

Yes, our project will assist handicapped individuals. Because many individuals in our nation are handicapped and unable to easily get to the hospital. They will be able to identify their ailment using a computer while at home in this situation.

#### **25. Conclusion:**

The system's major goal will be to predict Alzheimer's disease. The "MRI and Alzheimer's" dataset, released by the Open Access Series of Imaging Studies (OASIS) project, will be used to predict Alzheimer's disease or dementia in individuals. The missing values will be filled in and the dataset will be visualized. Some superfluous characteristics will be removed from the data during preprocessing. The values will be standardized to ensure that they can be readily included into machine learning models. The Adaboost classifier, extra-tree classifier, and SVM classifier will all be trained using the dataset. A confusion matrix will be utilized to assess the situation. Following that, we'll compare all three algorithms to see which one has the best accuracy. The system's dependability and performance will improve as a result of this. The ML method can assist the community understand the risk of dementia in patients by simply inputting MRI data. It is hoped that it may assist patients in receiving early dementia therapy and so enhance their quality of life.

#### **26. Bibliography:**

[1] <https://www.cdc.gov/aging/aginginfo/alzheimers.htm>



- [2] <https://www.nia.nih.gov/health/what-causes-alzheimers-disease>
- [3] [https://www.alz.org/alzheimers-dementia/research\\_progress/milestones](https://www.alz.org/alzheimers-dementia/research_progress/milestones)
- [4] <https://www.who.int/news-room/fact-sheets/detail/dementia>
- [5] <https://www.scitepress.org/Papers/2015/56152/56152.pdf>
- [6] <https://builtin.com/artificial-intelligence/machine-learning-healthcare>
- [7] <https://spd.group/machine-learning/machine-learning-in-healthcare/>
- [8] <https://www.hindawi.com/journals/jhe/2021/9917919/>
- [9] <https://builtin.com/data-science/tour-top-10-algorithms-machine-learning-newbies>
- [10] <https://www.dataquest.io/blog/top-10-machine-learning-algorithms-for-beginners/#>
- [11] <https://www.techiexpert.com/benefits-of-machine-learning-in-health-care/>
- [12] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3092139/>
- [13] <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01201-2>
- [14] [https://www.kaggle.com/jboysen/mri-and-alzheimers?](https://www.kaggle.com/jboysen/mri-and-alzheimers?select=oasis_longitudinal.csv)

[select=oasis\\_longitudinal.csv](https://www.kaggle.com/jboysen/mri-and-alzheimers?select=oasis_longitudinal.csv)

- [15] [https://www.researchgate.net/publication/50211052\\_A\\_Framework\\_for\\_Face\\_Recognition\\_Using\\_Adaptive\\_Binning\\_and\\_Adaboost\\_Techniques](https://www.researchgate.net/publication/50211052_A_Framework_for_Face_Recognition_Using_Adaptive_Binning_and_Adaboost_Techniques)
- [16] [https://en.wikipedia.org/wiki/AdaBoost#Real\\_AdaBoost](https://en.wikipedia.org/wiki/AdaBoost#Real_AdaBoost)
- [17] <https://www.geeksforgeeks.org/ml-extra-tree-classifier-for-feature-selection/>

[18] <https://www.geeksforgeeks.org/introduction-to-support-vector-machines-svm/>

[19] [https://en.wikipedia.org/wiki/Confusion\\_matrix](https://en.wikipedia.org/wiki/Confusion_matrix)

[20] <https://www.techopedia.com/definition/14650/data-preprocessing>

[21] <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>

[22] <https://appen.com/blog/training-data/>