FINAL PROJECT

ON

FACTORS INFLUENCING ALZHEIMER'S DISEASE

BY

Lakshmi Prabha Swaminathan

Raghuram Parthasarathy

Harikrishna Ramakrishnan

Table of Contents

Problem Statement

Alzheimer's is the most common form of dementia**,** a general term for memory loss and other intellectual abilities serious enough to interfere with daily life. Alzheimer's disease accounts for 60 to 80 percent of dementia cases. Alzheimer's has no current cure, but treatments for symptoms are available and research continues. Although current

Alzheimer's treatments cannot stop Alzheimer's from progressing, they can temporarily slow the worsening of dementia symptoms and improve quality of life for those with Alzheimer's and their caregivers. Today, there is a worldwide effort under way to find better ways to treat the disease, delay its onset, and prevent it from developing.

Overview of disease progression

The symptoms of Alzheimer's disease worsen over time, although the rate at which the disease progresses varies. On average, a person with Alzheimer's lives four to eight years after diagnosis, but can live as long as 20 years, depending on other factors.

Changes in the brain related to Alzheimer's begin years before any signs of the disease. This time period, which can last for years, is referred to as preclinical Alzheimer's disease.

The stages below provide an overall idea of how abilities change once symptoms appear and should only be used as a general guide. They are separated into three different categories: mild Alzheimer's disease, moderate Alzheimer's disease and severe Alzheimer's disease. Be aware that it may be difficult to place a person with Alzheimer's in a specific stage as stages may overlap.

1.2 Overview of disease progression

The symptoms of Alzheimer's disease worsen over time, although the rate at which the disease progresses varies. On average, a person with Alzheimer's lives four to eight years after diagnosis, but can live as long as 20 years, depending on other factors.

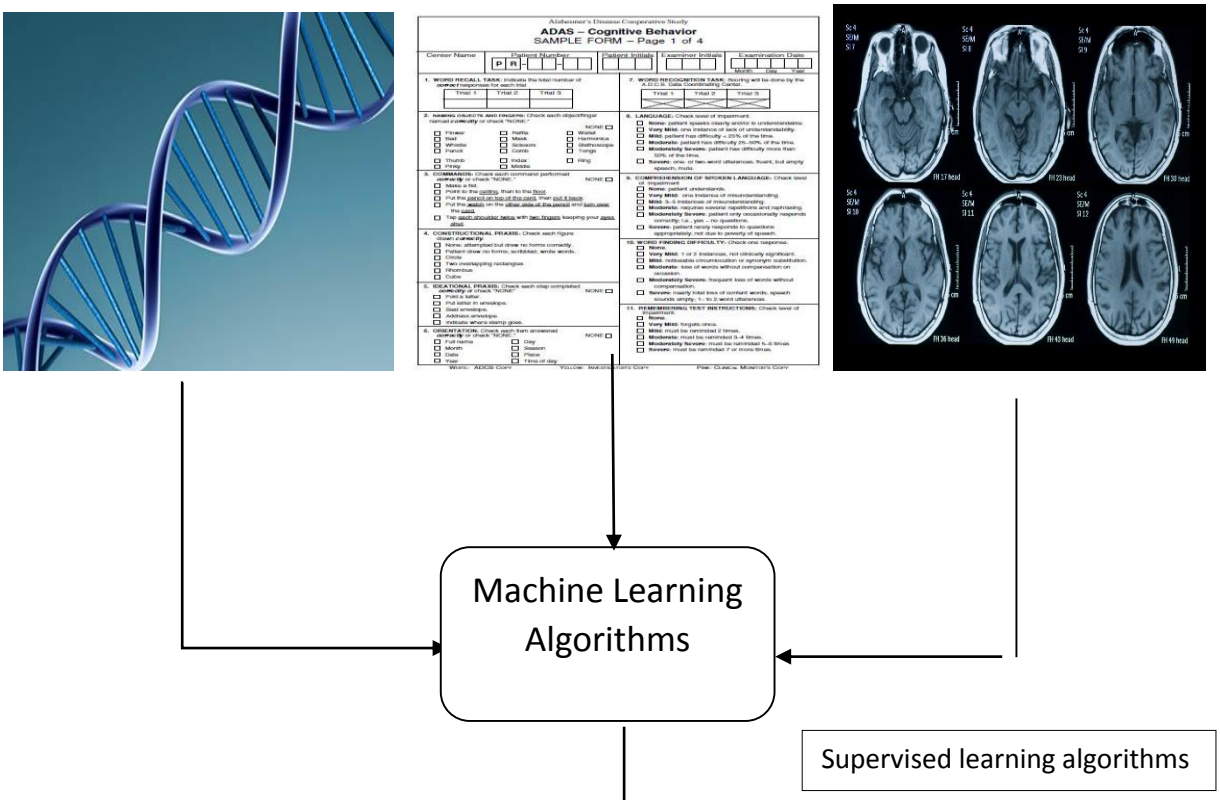
Changes in the brain related to Alzheimer's begin years before any signs of the disease. This time period, which can last for years, is referred to as preclinical Alzheimer's disease.

The stages below provide an overall idea of how abilities change once symptoms appear and should only be used as a general guide. They are

separated into three different categories: mild Alzheimer's disease, moderate Alzheimer's disease and severe Alzheimer's disease. Be aware that it may be difficult to place a person with Alzheimer's in a specific stage as stages may overlap.

1.3 Overall Project Architecture

Neurogenetic Dataset **Neuropsychological tests** **MRI Scans**



Project Objective

Our objective is to analyze cal tests, neurogenetic sets and brain scan images of patients and determine the presence of alzheimers in a patient. There are numerous neuropsychological tests taken by patients to determine the presence of Alzheimer. We determine which test is the best predictor of the disease. We perform principal component analysis on these datasets. We perform clustering in the neurogenetic dataset to determine the protein which contributes more to

the disease. We process MRI images of patients and analyze the hippocampus region of the brain to determine presence/absence of Alzheimer.

DataSets

Neuropsychological set

We consider five of the tests in our data analysis.

MMSE (Mini Mental State Examination)

- Orientation, word recall, attention and calculation, language abilities, visual construction
- Scores range from 0 to 30. Less than 26 indicates Alzheimer's

FAQ (Functional Activities Questionnaire)

- It is an informant (caregiver) based measure of functional abilities
- The informant provides details of target person on 10 complex, higher order activities.
- Levels of performance are assessed and ranges from dependence to independence.
- Response for 10 questions recorded. Each question can take 0-3 value.
- Scores range from 0 to 30. A cut-point of 9 (dependent in 3 or more activities) is recommended.
- • Dependent = 3 • Requires assistance = 2 • Has difficulty, but does by self = 1 • Normal = 0

NPI (Neuropsychiatric Inventory)

- Administered by clinician to the caregiver. He is asked a set of questions he needs to answer. He should be aware of the behavior of the patient.

- It assesses 10 or 12 behavioral domains in dementia.

GDS (Geriatric Depression scale)

- Response for 10 questions recorded. Each question can take 0-3 value.
- • 1–4 No cause for concern • 5–9 Strong probability of depression • 10+ Indicative of depression

MOCA (Montreal Cognitive Assessment)

- The MoCA test is a one-page 30-point test administered in approximately 10 minutes.
- Attention, concentration and working memory are evaluated using a sustained attention task (target detection using tapping; 1 point), a serial subtraction task (3 points), and digits forward and backward (1 point each). Language is assessed using a three-item confrontation naming task with low-familiarity animals (lion, camel, rhinoceros; 3 points), repetition of two syntactically complex sentences (2 points), and the aforementioned fluency task. Finally, orientation to time and place is evaluated (6 points).

Neurogenetic set

Translocase of outer mitochondrial membrane 40 homolog (yeast), also known as TOMM40, is a protein which in humans is encoded by the TOMM40 gene.

Clinical significance

In humans, certain alleles of this gene have been statistically associated with an increased risk of developing late-onset Alzheimer's disease. One

study has found that TOMM40 risk alleles appears twice as often in people with Alzheimer's disease than those without it.

TOMM40 Allele 1 and TOMM40 Allele 2 was clustered to identify better range distribution.

MRI Images (Brain Scans)

When a person is affected by Alzheimer, the hippocampus region of the brain degenerates. Alzheimer patients suffer from memory loss because the hippocampus region of the brain keeps track of all memories. In the neuroimaging dataset, we identify the hippocampus region of the brain and indirectly measure its degradation.

Data Analysis

Neuropsychological Test – PCA

When a person is affected by Alzheimer, the hippocampus region of the brain degenerates. Alzheimer patients suffer from memory loss because the hippocampus region of the brain keeps track of all memories. In the neuroimaging dataset, we identify the hippocampus region of the brain and indirectly measure its degradation.

We perform this to determine the significant test of the 5 tests

The tests we consider are

- NPI
- FAQ
- MMSE
- CDR
- MOCR
- GDS

We need to aggregate the scores for each of the tests before we perform principal component analysis.

Net score over a period of time = (Score 1 + Score 2+…+Score n)/Day Difference

```
dbSendQuery(mydb, "select m.RID, m.comp_MMSE_score, n.comp_FAQ_total,o.comp_NPI_total
             p.comp_GDS_total, s.comp_CDR_total,t.comp_MOCA_total from

            (select a.RID,MMSCORE/daydiff as comp_MMSE_score from
            (select RID,sum(MMSCORE) as MMSCORE from MMSE_clean group by RID) a
            inner join
            (select RID,datediff(max(update_stamp),min(update_stamp)) as daydiff
            from MMSE_clean group by RID) b on a.RID=b.RID where
            (MMSCORE/daydiff) is not null) m
```

We perform this for each of the test and join the aggregated data for all the tests

This is given as input for principal component analysis.

```
data = fetch(rs, n=-1)

psychpca = data[,2:6]
#Omitting NA and performing principal component analysis
pca_comp <- prcomp(na.omit(psychpca), scale. = TRUE)
```

Neurogenetic Dataset (Clustering)

We perform k means clustering on the neurogenetic dataset to determine the distribution of protein in the two genes (TOMM40_ADNI1 and TOMM40_ADNI2)

Result:

We determine that ADNI2 is the dominant protein after performing k means clustering.

Neuroimaging – Computing size difference

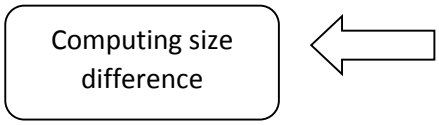
Calculating hippocampus size difference

Steps involved:

1. Neuroimage segmentation

2. Laplacian transform to normalize image
3. Detection of edges
4. Detection of corners with the edges already detected
5. Calculation of size difference

1. Image segmentation 2. Laplacian Transform 3. Edge detection



Corners detection

File processing for size difference computation

1. The datasets of the MRI scans have been downloaded from http://www.oasis-brains.org/(http://www.oasis-brains.org/app/template/Tools.vm;jsessionid=35D9DD2D26127CF9EC731915BD6B698E#ftp)
2. We download the cross sectional view dataset (total of 12 disks)

3. We write a python code to extract the required processed MRI images to another folder so that we can perform image processing on them.

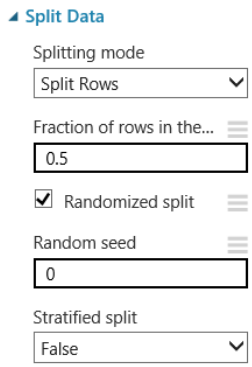
```
frame = cv2.imread(wname,cv2.IMREAD_COLOR)
laplacian = cv2.Laplacian(frame,cv2.CV_64F)
edges = cv2.Canny(frame,100,150)
cv2.imshow('original',frame)
```

   Laplacian is used for image normalization and Canny is used for edge detection.

4. We extract only those images where a patient has 2 MRI scans (to compute size difference)
5. We run the python code that performs the image processing to store the size difference in the database.

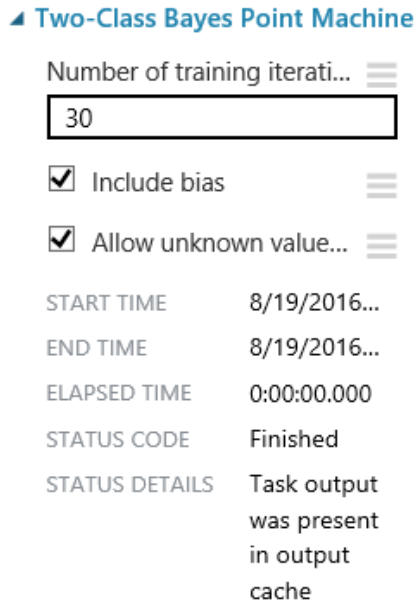
AzureML Studio:

The *Split* module in Azure ML is used to split the data into train and test sets where the split is stratified. The stratification will maintain the class ratios into the two output groups. We use the first 50% of the dataset for training and the remaining 50% for testing the performance of the trained model.

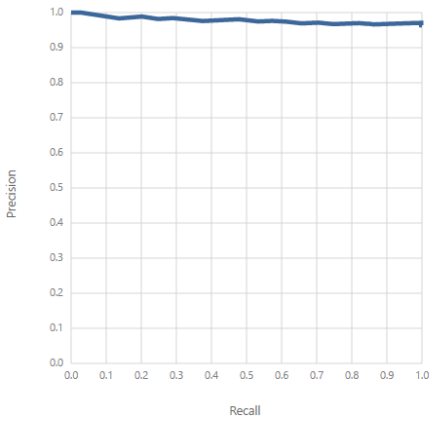


Two class Bayesian Point Machine

For the two class Bayesian classification, we set the number of iterations to 30.

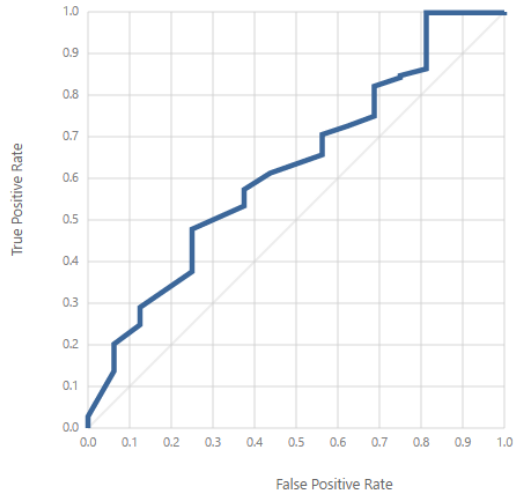


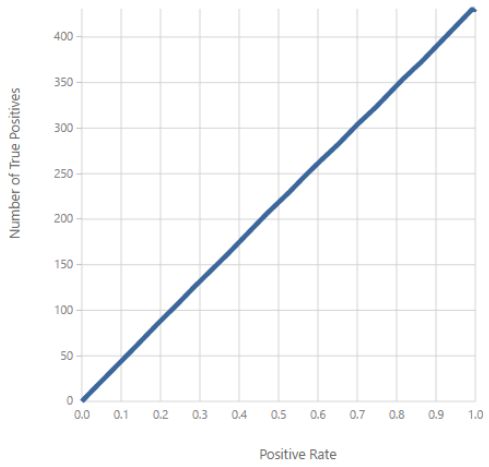
Precision Curve



ROC Curve

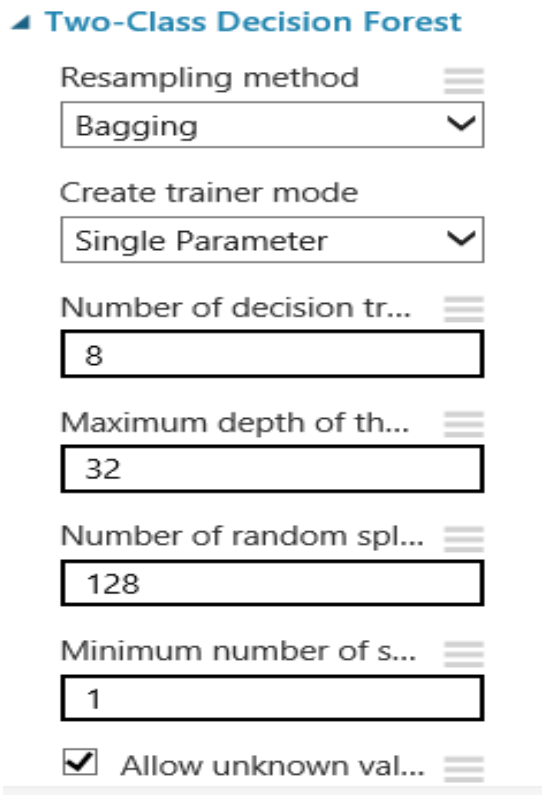
LIFT Curve





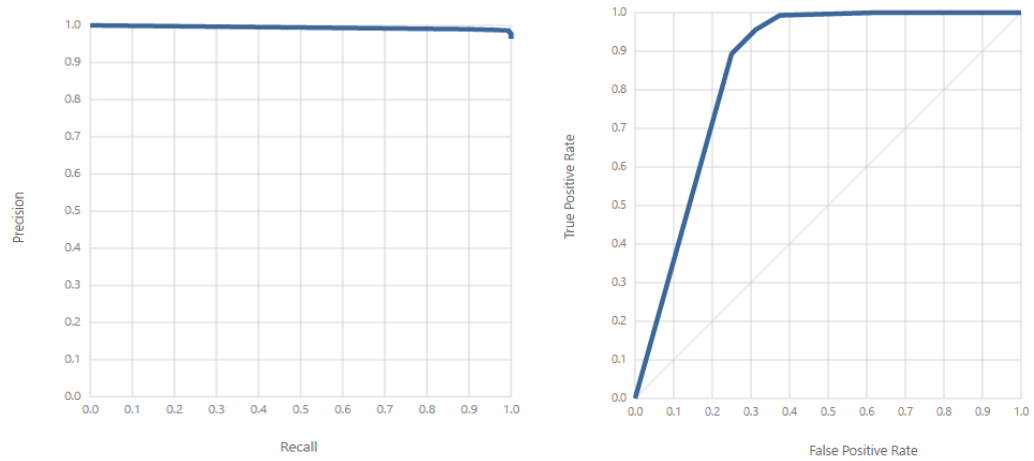
Two class decision forest

We set the number of decision trees to 8.



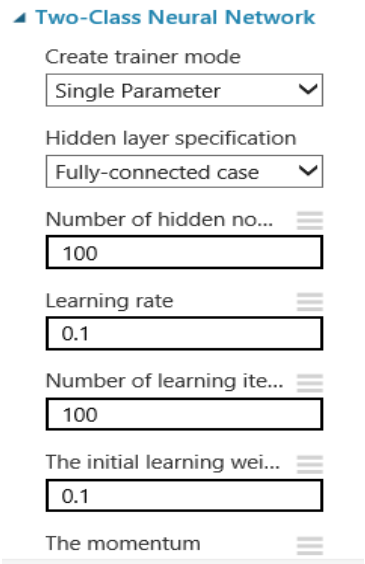
Random Forest

Precision Curve ROC Curve



Two Class Neural Network

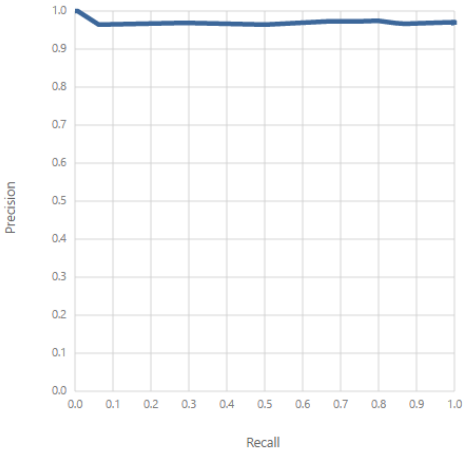
The number of hidden nodes is set to 100



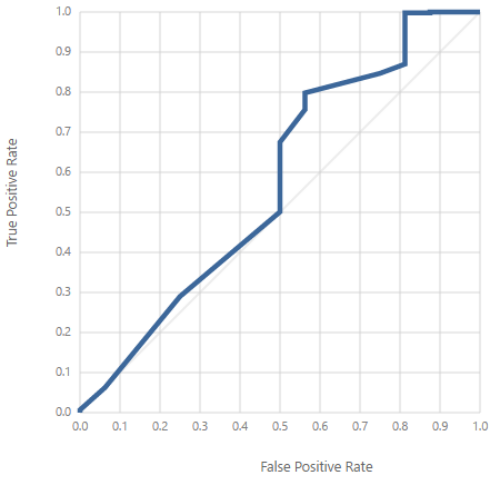
Train prediction model

To train the model, we connected the text features created in the previous steps (the training data) to the***Train Model** module. Microsoft Azure Machine Learning Studio supports a number of learning algorithms but we select SVM for illustration. The Alzheimer_flag is used a column selector

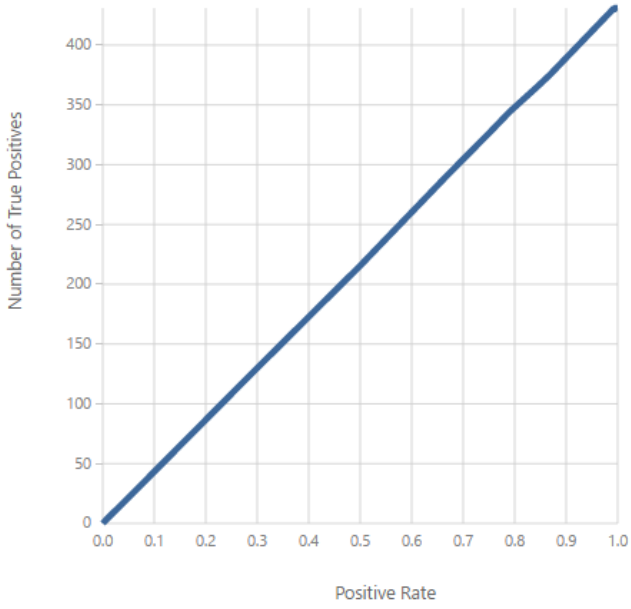
Precision Curve



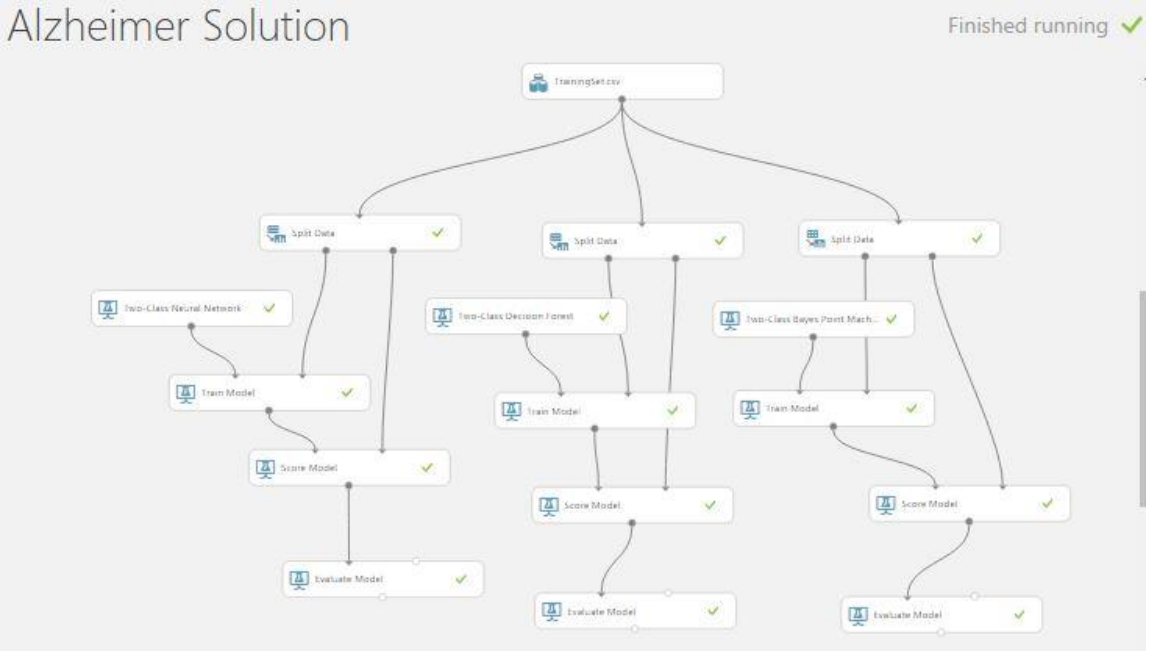
ROC Curve



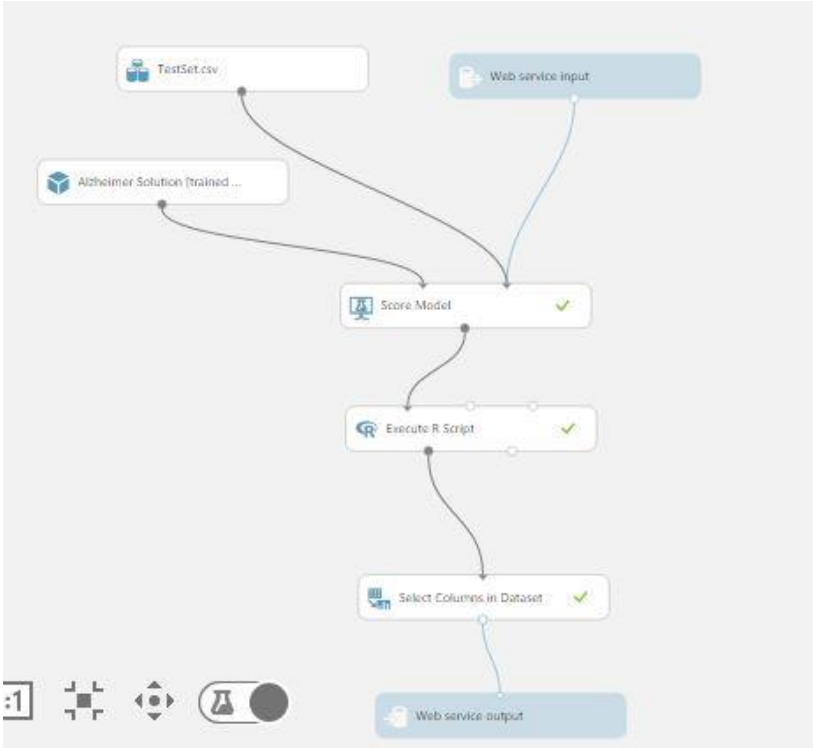
Lift Curve



Experiment



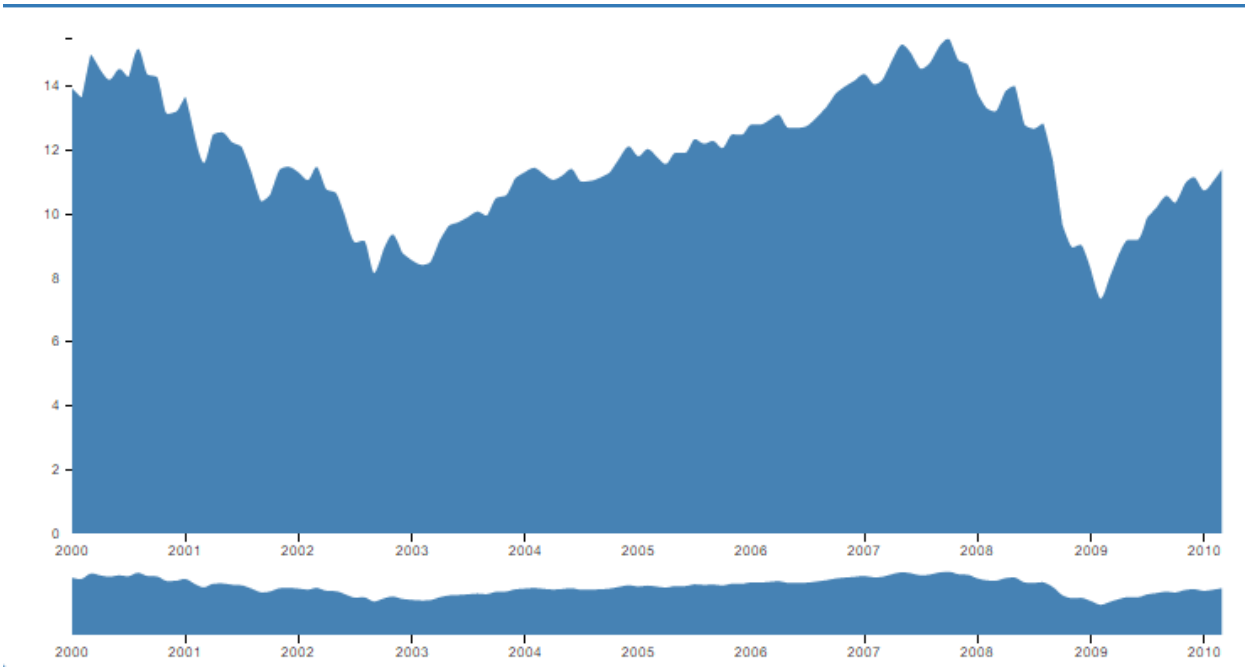
Predictive Model



Visualization:

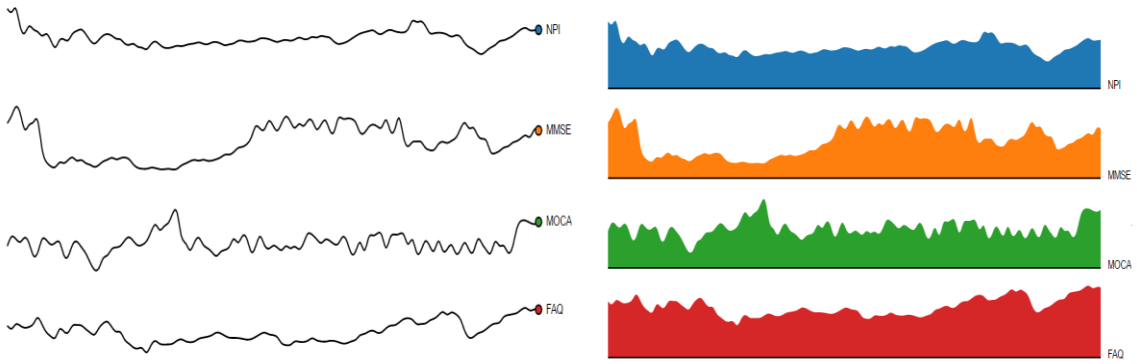
Graph1:

The below graph shows the combined average of all scores from the year 2000 to 2010. The special feature of this graph is that the user will be able to select a particular month/year by snipping the required portion the user wants to visualize.

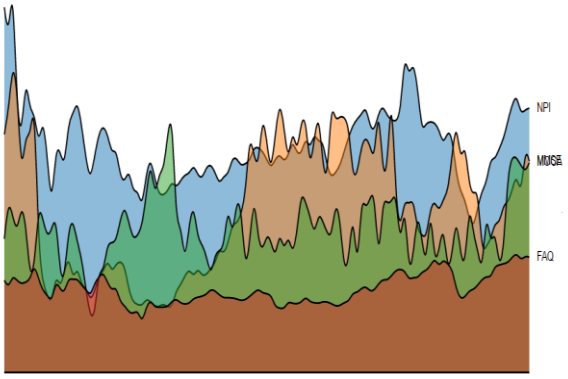


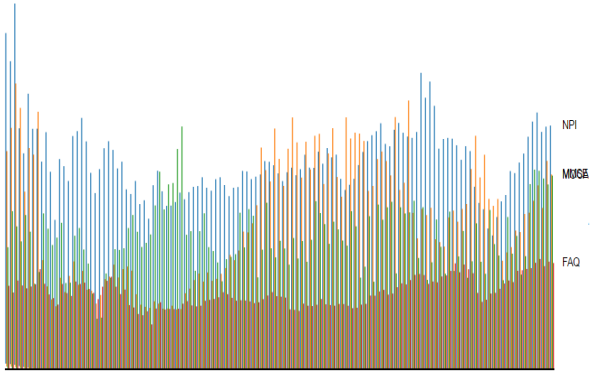
Graph2:

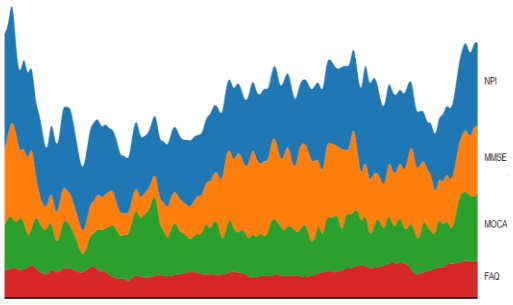
This graph shows the combined average of all scores from the year 2000 to 2010.

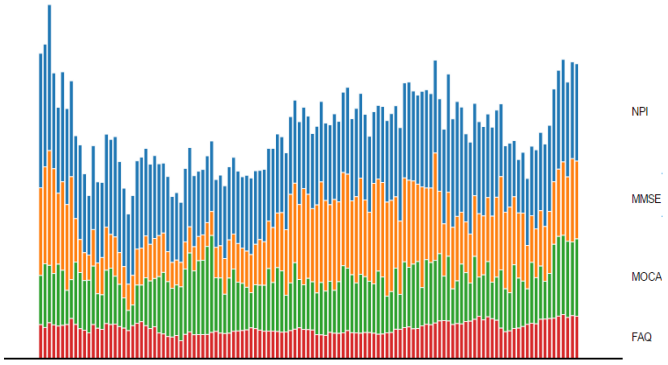


The special feature of this graph is that the user will be able to visualize dynamic charts with different formats such as Stacks, line graph, Pie chart and Bar diagram



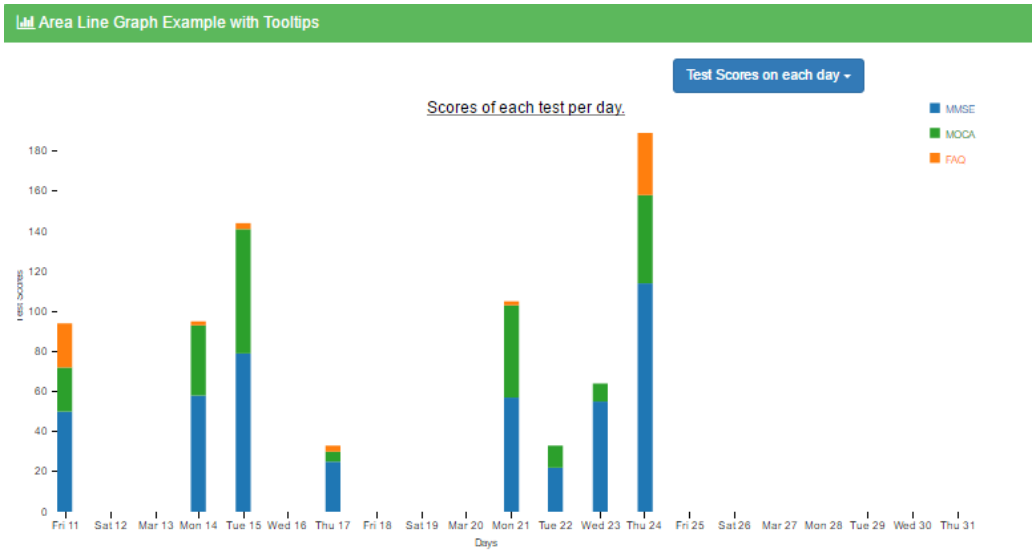






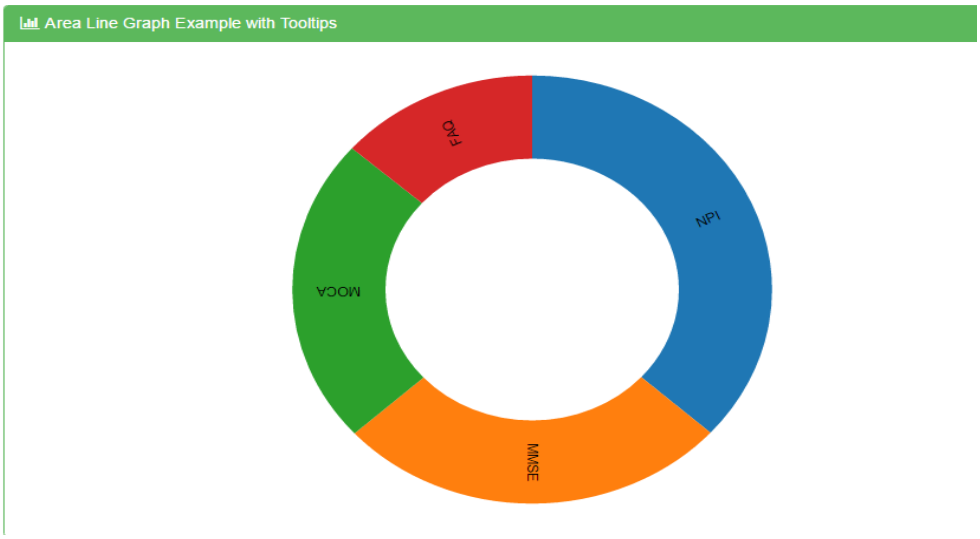
Graph3:

This graph shows the test scores for three different tests (MMSE, MOCA, FAQ) for a selected day in a month. The special feature of this graph is that the user will be able to view the scores of individual tests taken by different patients during the different dates.



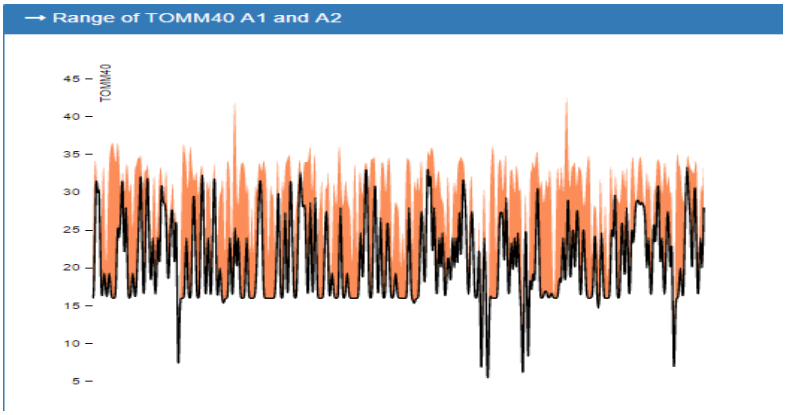
Graph 4:

The Donut represents the *average* price (or sum) during this time period.



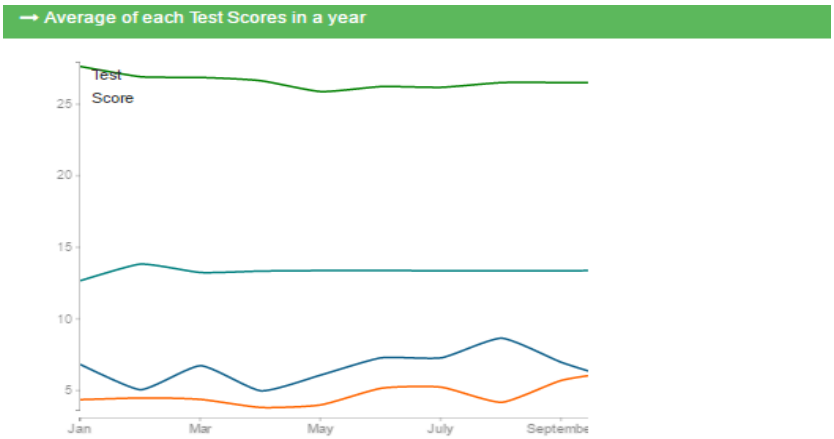
Graph 5 :

The below graph displays the range between two **Translocase of outer mitochondrial membrane (TOMM40 A1 and A2)** where A2 is the dominant part used for calculating the size difference inside the brain.



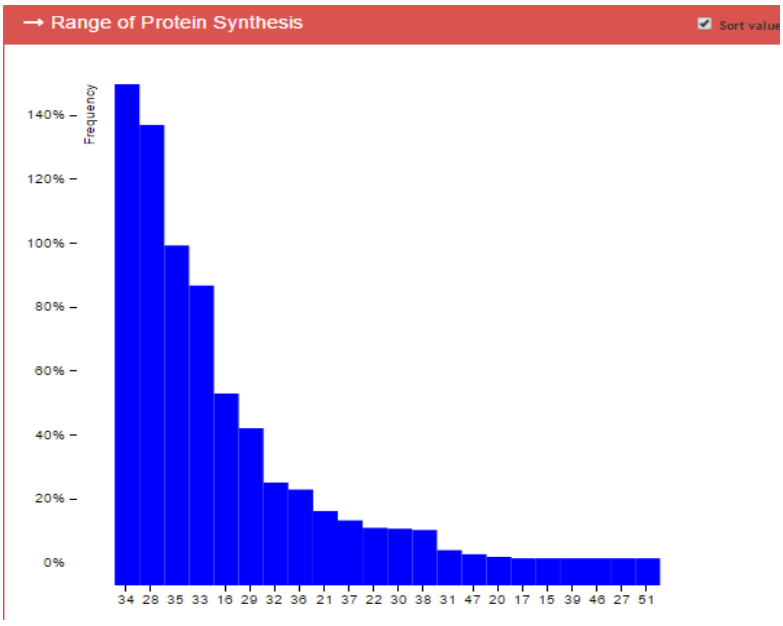
Graph 6:

This graph shows the average score of each test performed during the year 2011

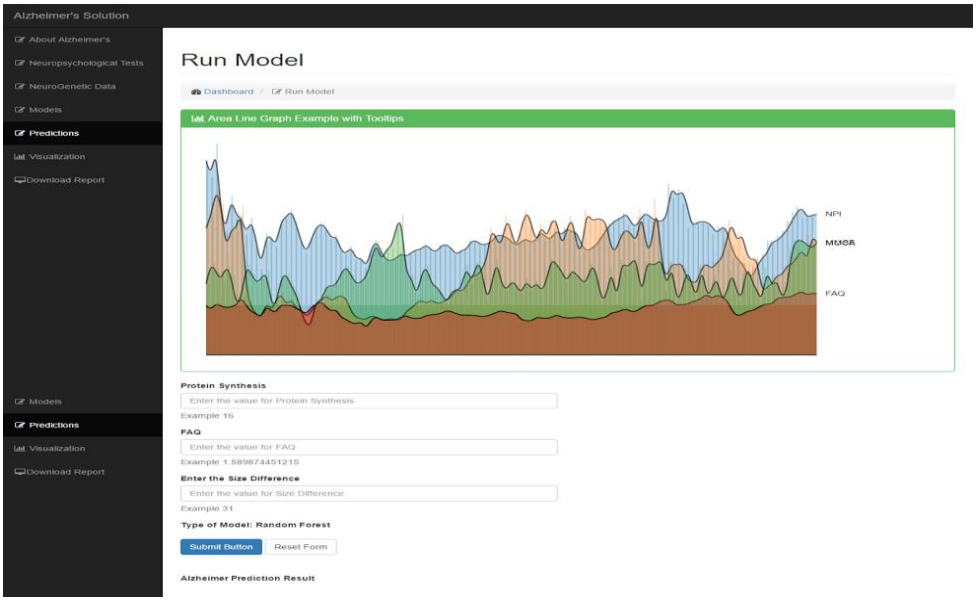


Graph 7:

The below graph shows the frequency, i.e. number of occurrences of protein synthesis values, of different patients taken during a 2000 - 2010

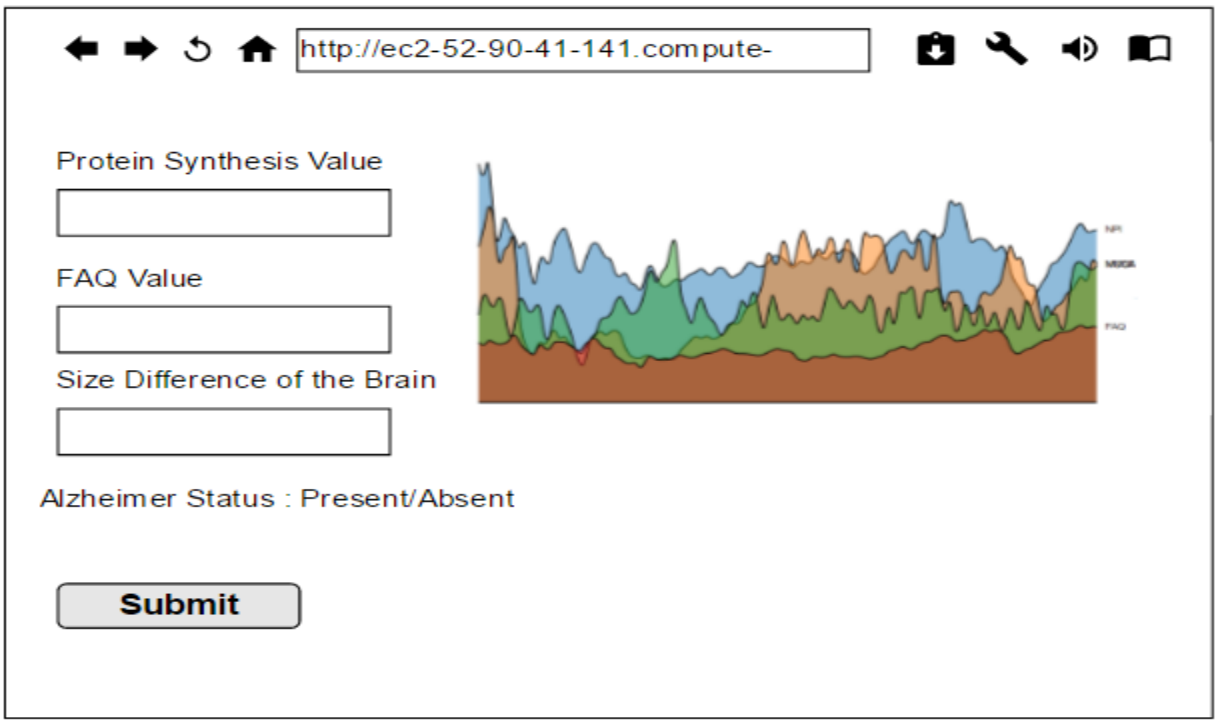


User Interface Screen:

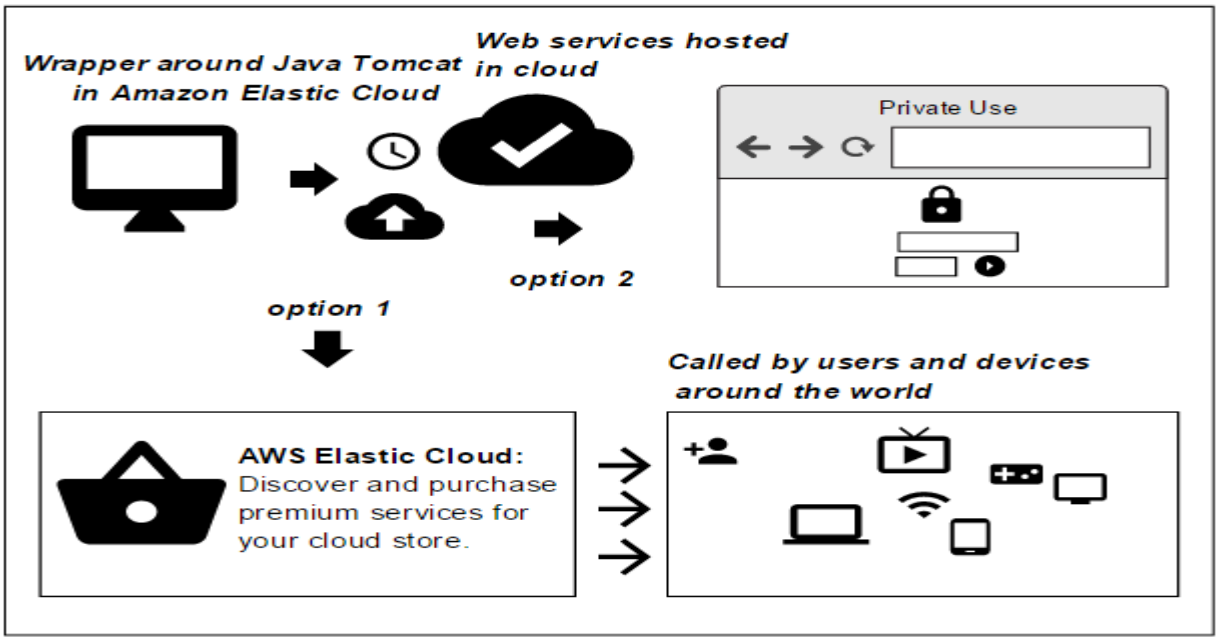


Moque Ups:

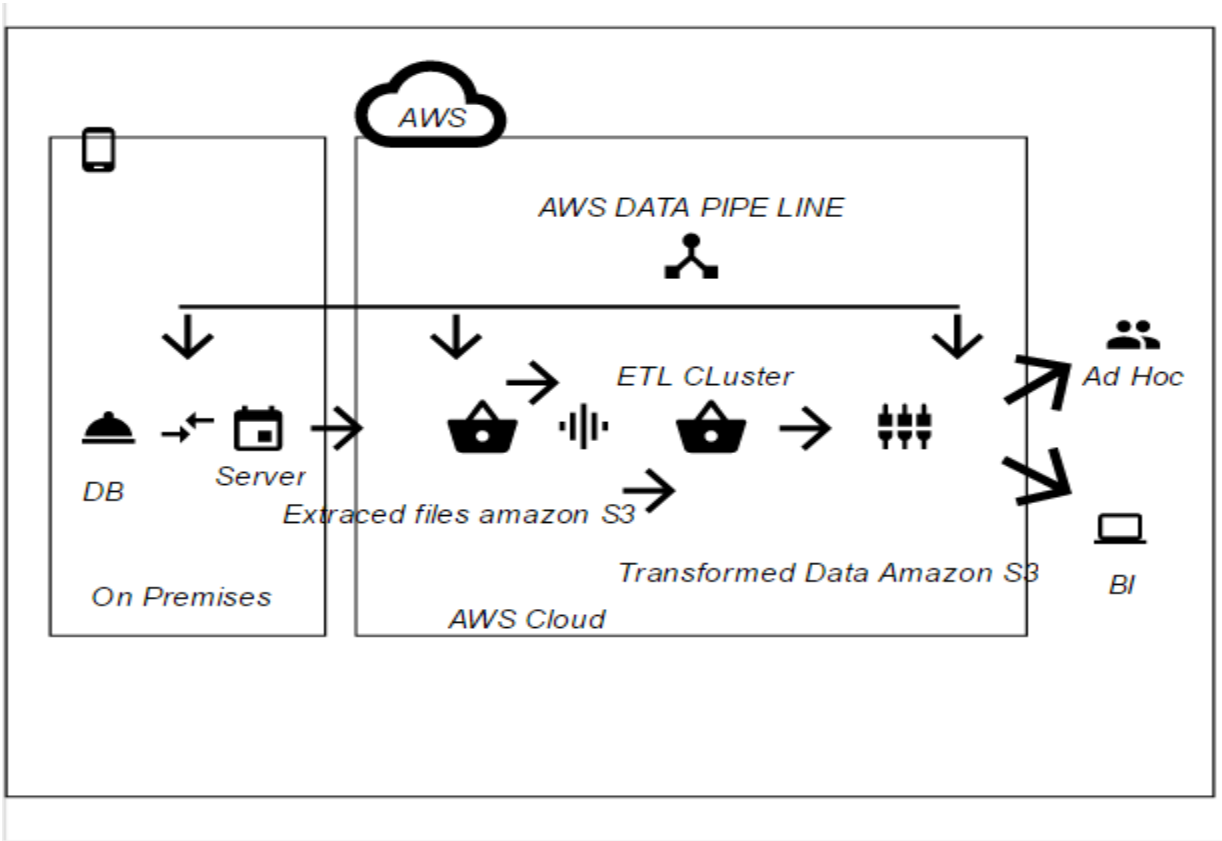
Moque Up 1:



Moque Up 2:



Moque Up 3:



Conclusion

The system will assist physicians to predict Alzheimer's accurately. Overall benefits achieved from the system include:

- Isolating significant features from least significant tests
- Integrating these significant isolated features
- Interpreting Alzheimer's in a patient

Limitations of the system:

- Severity of the disease (Mild or moderate or severe) cannot be assessed by the system
- Limited spatial resolution of MRI scans might result in slight deviation of size differences observed