## 

## 

ALZHEIMER’S DISEASE DETECTION AND CLASSIFICATION

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## **Team Members:**

## Tamilarasee Sethuraj – A20553416

* Krishnavardhini Kittusamy - A20502176

## **Introduction**

Alzheimer’s disease (AD) is the most common type of Dementia, a [neurodegenerative brain disease](https://www.sciencedirect.com/topics/medicine-and-dentistry/degenerative-disease) that can significantly affect the quality of life of older individuals. This is a long-lasting problem that slowly reduces and ultimately kills one's thinking abilities and leads to a huge loss of cognitive and mental strength, behavioral problems as well as language difficulties.

The commonly used method is to predict the disease based on Magnetic Resonance Imaging (MRI) for the purpose of early intervention. Although there is no cure for Alzheimer's, there are medications and interventions to manage symptoms and slow down the progression to some extent. Early treatment proves to be an effective way of preventing its deterioration.

The objective of the project is to preprocess the MRI images, explore and use machine learning algorithms to classify brain MRI images into different stages of Alzheimer's disease. We will evaluate the trained models using appropriate evaluation metrics and compare their performance to identify the most effective model with highest possible accuracy and visualize the results.

## **Problem Description**

The focus of the problem is to identify subtle changes in cognitive function that may indicate the severity of AD which would help providers for making informed decisions about patient care and early treatment.

Many scientists have made sincere efforts to discover a variety of techniques to detect Alzheimer's using MRI data. Those techniques include the extraction of discriminative features from a large set of features, and selecting efficient classification models from machine techniques. While most of the early techniques deploy binary classification of finding whether the given MRI image is subject to AD or not, there are some recent techniques which focus on finding the current stage/severity of the disease such as very mildly demented, mildly demented, moderately demented.

Some of the difficulties that scientists have faced in implementing the models are extracting clear image features that show small variants of brain cells changes, using complex models restricts the ability to provide relevant information based on which the decision was made (classification). It is important to know the logic/root of the decision as it involves patient care and hence, we may need to rely on simple explainable models which may not have much scope for accuracy or improvement.

## **Data Description**

We have used MRI images as input for our classification problem. The images are collected from the data set “Alzheimer MRI Dataset” from Kaggle which was collected from several websites/hospitals/public repositories.

## The dataset comprises 6400 preprocessed MRI images, all resized to 128 x 128 pixels, depicting various stages of Alzheimer's disease progression. There are a total of 6400 images.

## They are categorized into four classes based on the severity of dementia:

## Class 0: Non-Demented (3200 images)

## Class 1: Very Mild Demented (2240 images)

## Class 2: Mild Demented (896 images)

## Class 3: Moderate Demented (64 images)

## **Completed Tasks**

We have worked on Data preparation for training the models and implemented Logistic Regression and Random forests algorithms to classify the Alzheimer’s disease and observed the accuracy of the classification.

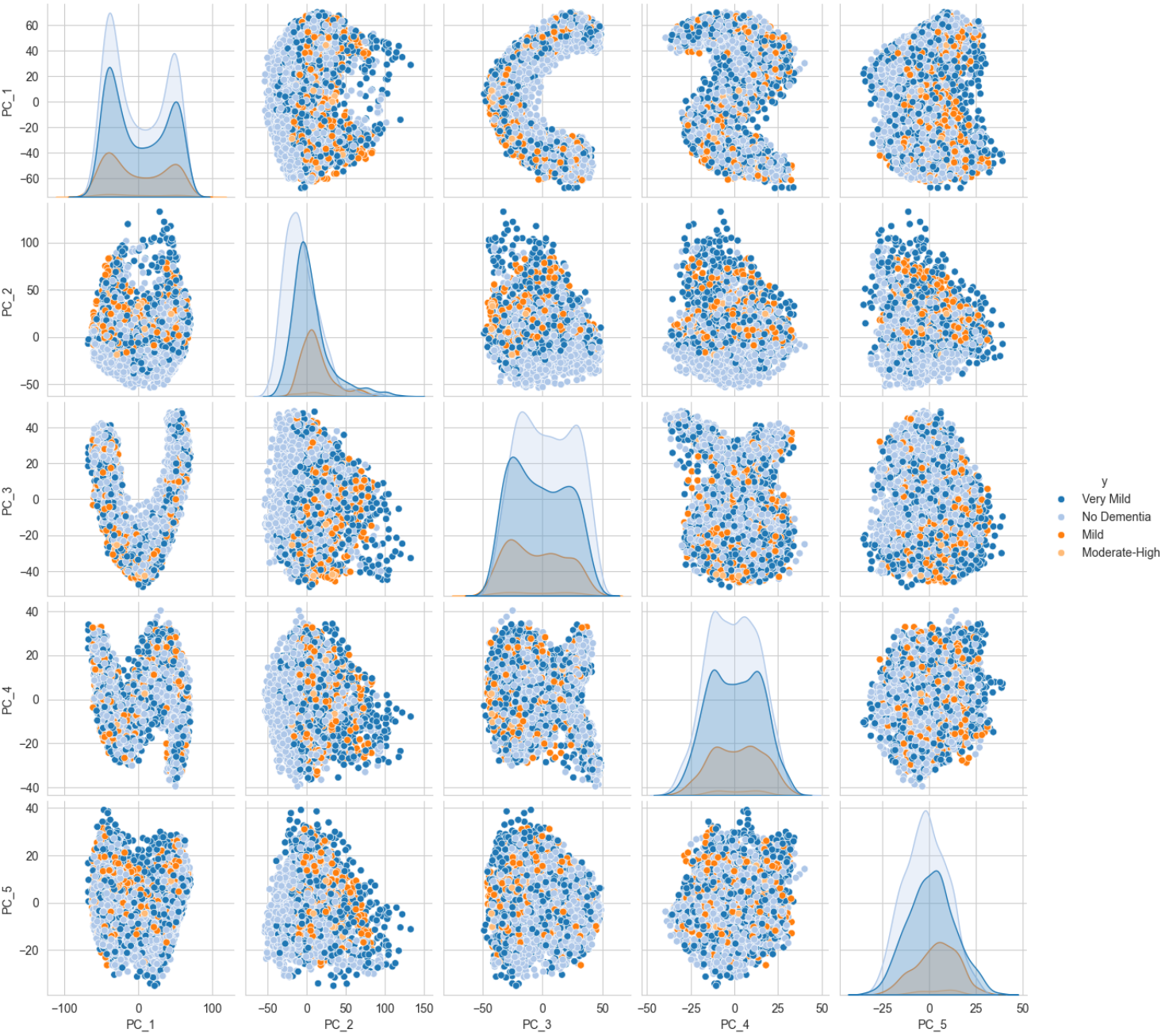
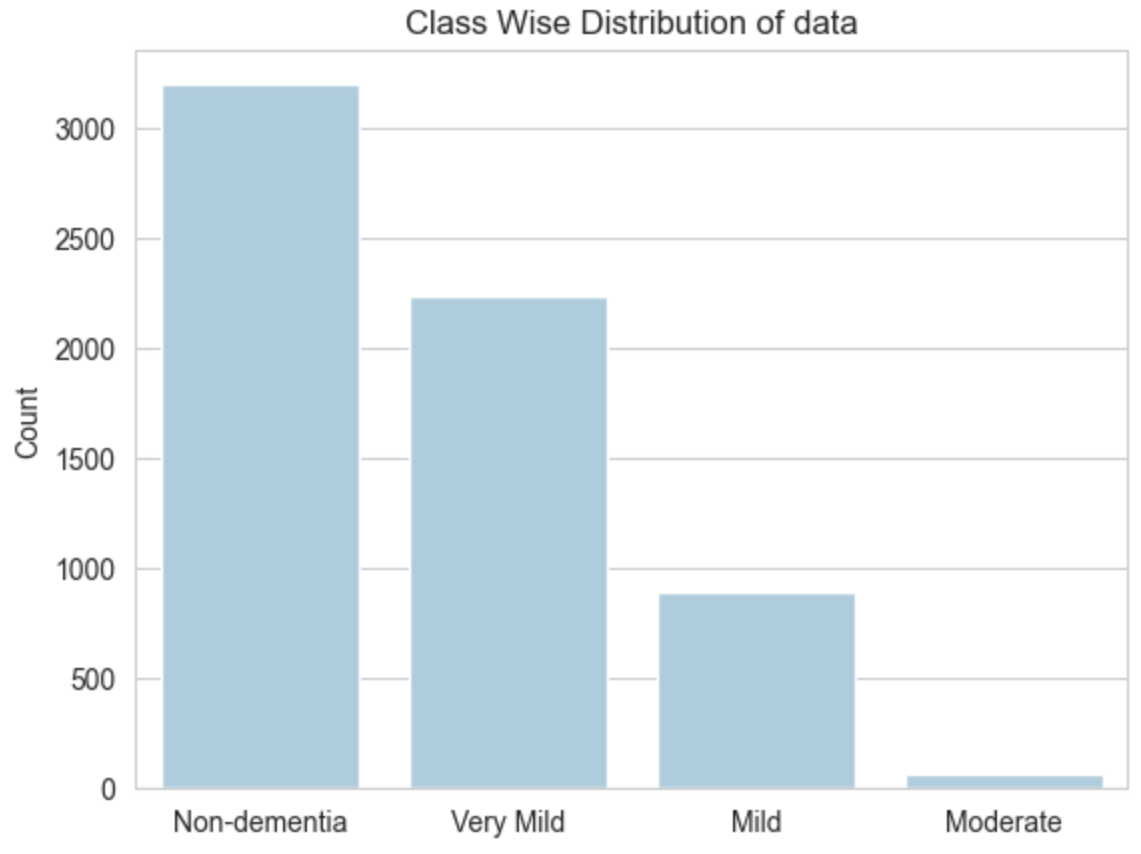
### **Data Preparation:**

We have cleaned and processed the collected data for the implementation. We performed several preprocessing and augmentation steps on the dataset to enhance data quality and improve the model's robustness. The data is preprocessed by scaling it and then applying PCA to reduce its dimensionality, ensuring that the transformed data retains most of its variability.

Some of the key steps followed are as below:

* Image Preprocessing - All images are resized to a standard size of 60 x 60 pixels. The images are flattened to an one dimensional array and the y labels are added to create a dataframe for each class.
* Dataset Splitting - The processed dataset is divided into train, validation, and test sets.
* Data Augmentation - Augmentation techniques like rescaling, shearing, and zooming are applied to the training data to increase diversity and improve generalization.
* Data Normalization - Validation and test data are rescaled to ensure normalization.
* Directory Setup - Directories are organized to specify the locations of split images for the train, validation, and test sets.
* ImageDataGenerators Usage - Keras ImageDataGenerator is utilized to create batches of augmented images for training and normalized images for validation and testing.
* Class Mode Configuration - The class mode is set to 'categorical' to support multi-class classification tasks.

Below plots shows the distribution of the data across classes



These steps collectively prepare the dataset for training and evaluating deep learning models aimed at classifying Alzheimer's disease using MRI images of the brain.

### **Logistic Regression:**

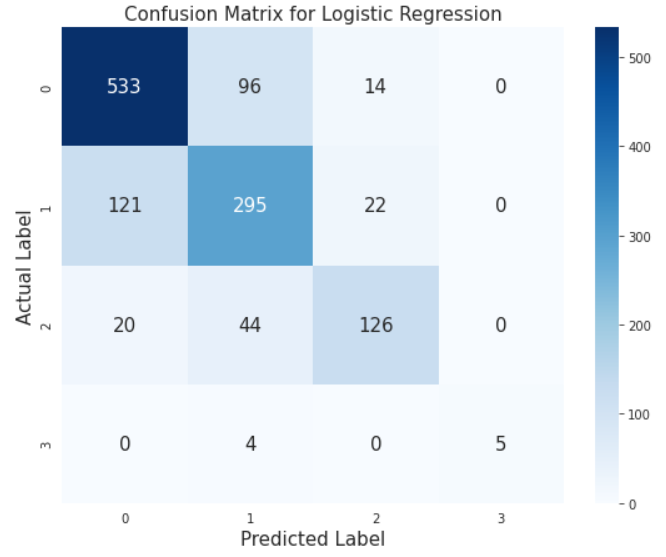
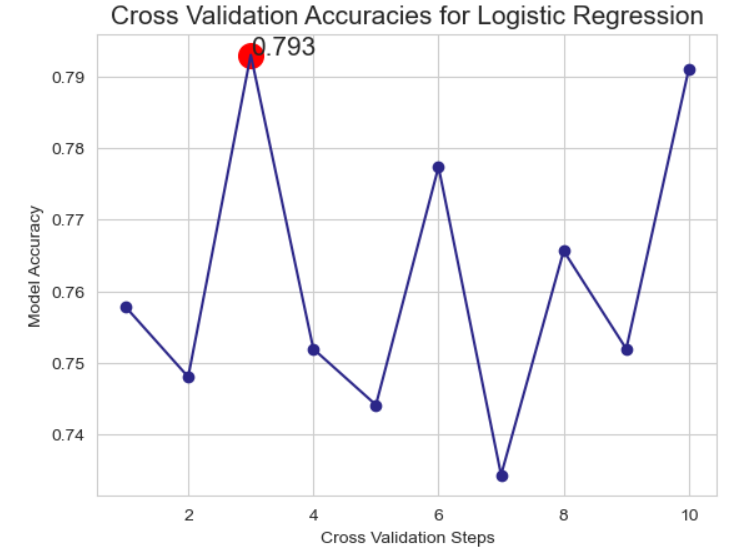
In our project we have used Logistic regression as one of the classification algorithms to model the data and used multi classification as we have more than one class of the data. Logistic regression predicts the probabilities of each possible class. The model outputs a vector of probabilities, one for each class. To ensure that the predicted probabilities sum up to 1, the Softmax function is used. It exponentiates each raw output and then normalizes the results. It transforms the raw outputs of the model into probabilities.

We work with MRI images to identify patterns associated with AD. Feature selection techniques are employed to reduce dimensionality and enhance classification performance. We set up a Logistic regression model with specified parameters . It then defines a grid of hyperparameters, for hyperparameter tuning using grid search cross-validation. In training, the model is optimized using a loss function such as cross-entropy loss with maximum likelihood of the correct class label given the input features.

We have used 'l2' (ridge) regularization terms to penalize the large coefficients and fine tuned the regularization parameter for a set of iterations to prevent overfitting. Grid search cross-validation, and various evaluation metrics tools are used to verify the performance of various combinations of the hyper parameters in the model. Regularization parameter value of 0.001 and the Newton-CG optimization algorithm are used to find the appropriate coefficients of the model to arrive at the better accuracy of 79%.

From the confusion matrix and other evaluation metrics , we see that Class 0 and Class 3 have been relatively well-classified, with high precision, recall, and F1-score while Class 1 and Class 2 show moderate performance.

Below are the graphs for Cross validation and Confusion Matrix for the logistic regression



### **Random Forest**

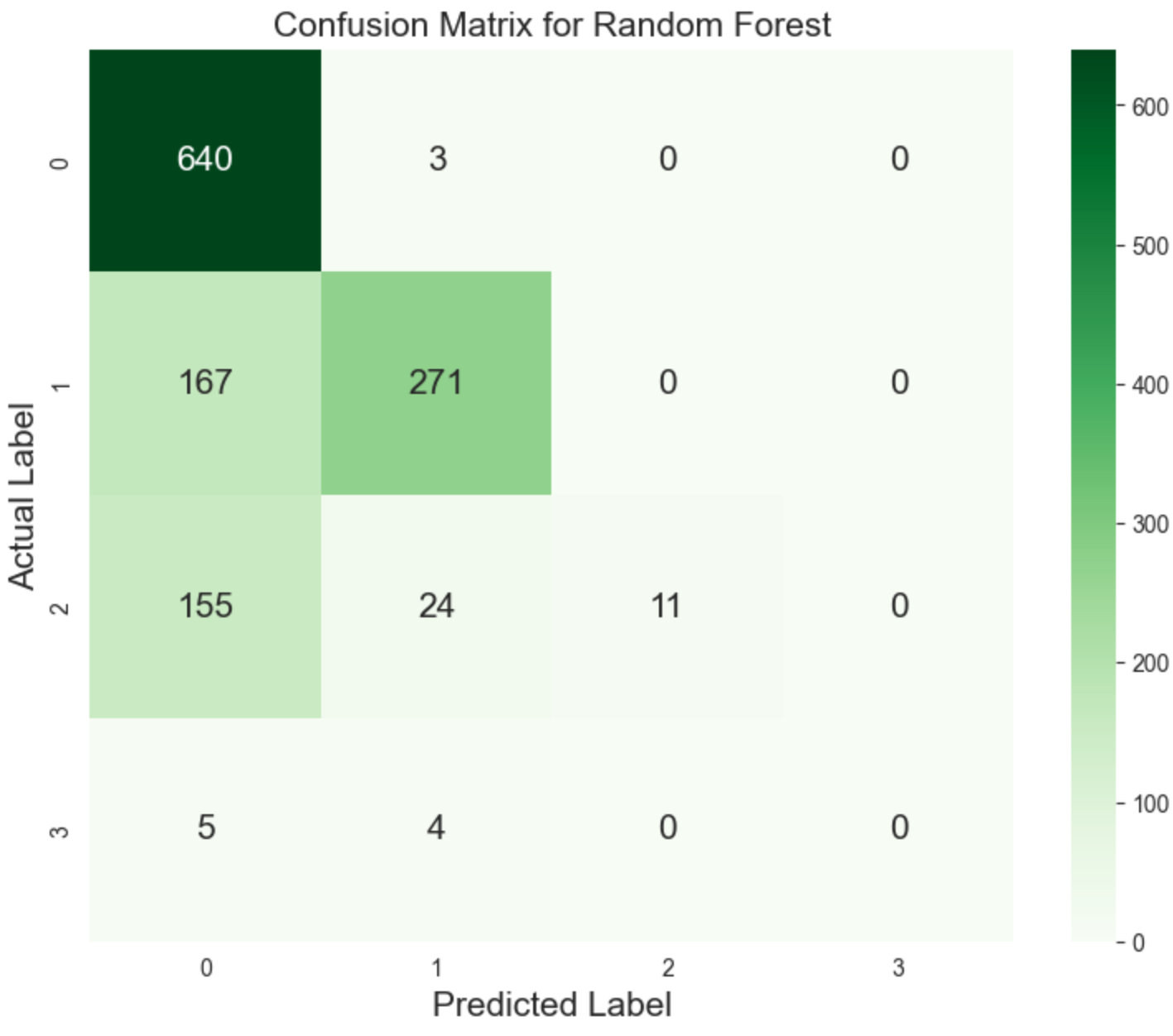
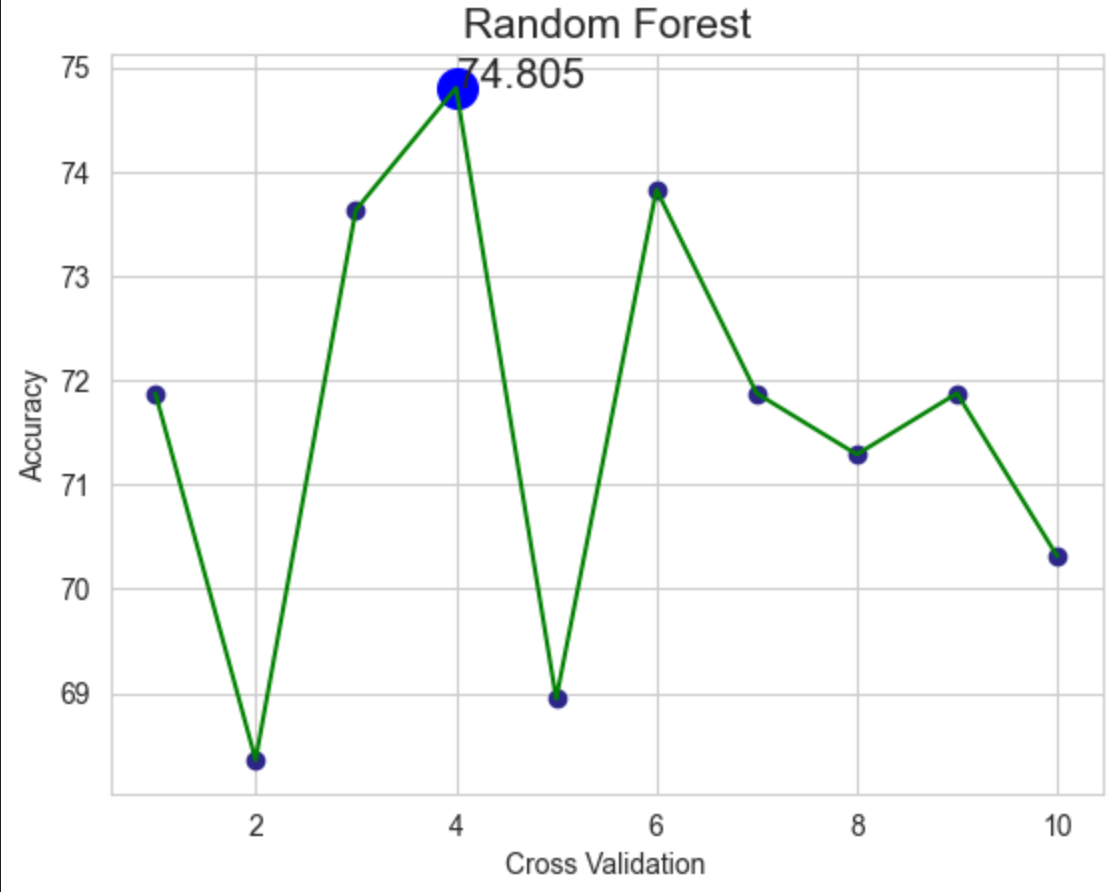
The Random Forest algorithm is a popular machine learning technique used for classification tasks. Random forest is based on the concept of decision trees. A decision tree is a method in which the internal nodes represent the “question logic” and the node branches based on the decisions it can take, thereby reaching the leaf node that denotes the class label/prediction.

Random forest takes a random subset of training data each time and builds multiple decision trees.At each node of the decision tree, only a random subset of features is considered for splitting. Random forest prediction is based on the number of decision trees which predict the same class.

The two main hyper parameters of the algorithm are the number of trees in the forest and the number of features that we choose to extract. We have tried a number of combinations to identify the efficient parameters for our dataset.

The square root of the number of features and 1000 decision trees finds the optimal balance between model performance and computational efficiency which is around 74.8%.

We plotted the confusion matrix to visualize the impact of the model on the dataset. The model performs reasonably well for class 0, with relatively high precision, recall, and F1-score. However, it performs less effectively for class 1, with lower precision, recall, and F1-score compared to class 0. Class 1 seems to have been mislabeled more compared to other classes. Classes 2 and 3 also show varying levels of performance, with relatively lower support and precision.



## **Future Tasks**

We have worked on the traditional algorithms logistic regression and random forest and found the accuracy to be 79% and 75% respectively.

We will explore SVM and CNN models for our dataset and compare the performance of the approaches and identify the most efficient method. We will try to explore other features that could be imparted into the method to enhance the existing method and to ensure high accuracy and reliability.

We aim to update parameters of the efficient model to allow for traceability of decisions, ensuring it is transparent and suitable for clinical purposes. Overall, the goal of the proposed system is to improve Alzheimer's Disease classification, making it more efficient and accurate for easier integration into clinical practices, ultimately benefiting patient care.

## **References**

* A deep learning model for Alzheimer’s disease diagnosis based on patient clinical records<https://www.sciencedirect.com/science/article/abs/pii/S0010482523012799>
* [Bharath M](https://ieeexplore.ieee.org/author/37088534615); [Gowtham S](https://ieeexplore.ieee.org/author/37089590515) ; [Rohini B R](https://ieeexplore.ieee.org/author/37089953400) - Predicting Alzheimer’s Disease Progression through Machine Learning Algorithms - [2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET)](https://ieeexplore.ieee.org/xpl/conhome/10419871/proceeding)
* <https://www.researchgate.net/publication/322584207_A_deep_CNN_based_multi-class_classification_of_Alzheimer's_disease_using_MRI>
* Comparative study of detection and classification of Alzheimer's disease using Hybrid model and CNN -<https://ieeexplore.ieee.org/document/9688082> - [2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)](https://ieeexplore.ieee.org/xpl/conhome/9687822/proceeding)
* https://www.researchgate.net/publication/378090880\_Application\_and\_comparison\_of\_K \_means\_and\_PCA\_based\_segmentation\_models\_for\_Alzheimer\_disease\_detection\_using\_MRI
* Eke, Chima S., et al. "Early detection of Alzheimer's disease with blood plasma proteins using support vector machines." IEEE journal of biomedical and health informatics 25.1 (2020)