International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 13, No.1 (Jan-2023)

http://dx.doi.org/10.12785/ijcds/130121

AD ALERT Classification of Alzheimer disease with Traditional and Deep Network Models

Vunnam Asha latha [1] **and Anupama Namburu**[2]

[1,2]*School of Computer Science Engineering, VIT-AP University, India*

Received 3 Feb. 2022, Revised 3 Jul. 2022, Accepted 12 Jan. 2023, Published 31 Jan. 2023

Abstract: The most common cause of dementia is Alzheimer's disease(AD). Alzheimer's disease has a slow rate of advancement, which gives patients the chance to receive early treatment through regular testing. However, due to their high price and restricted availability, current clinical diagnostic imaging techniques do not satisfy the specific needs for screening methods. AD ALERT aims to address the problem by automating the detection of the Alzheimer's with machine learning techniques. In this paper, the Magnetic Resonance Image (MRI) data is extracted, and feature selection based on random forest is used to select top 30% important and useful features among the total features for the analysis. The deep network models were proposed to classify the patients to AD using these selected features based on random forest. The traditional classification techniques namely K Nearest Neighbour (K-NN), decision tree, Stochastic Gradient Descent (SGD) , and Support Vector Classifier (SVC) were also implemented using the same feature selection model to compare the performance of the deep network models. The deep network models proved to be outperforming the considered model with an accuracy of 0.98%.

Keywords: Alzheimer classification, Alzheimer's dementia, Classification, Deep learning, Machine learning, MRI data extraction

1. INTRODUCTION

Alzheimer's disease has become one of the world's main diseases, according to the World Health Organization (WHO). Alzheimer's disease affects people all around the world in various ways. Dementia is estimated to grow to 82 million persons in 2030 and 152 million by 2050 [1]. Alzheimer's disease is characterised by memory loss [2]. Patients eventually lose track of their personal information, such as their name, age, and address [3].They even lose track of their family members and relatives at some point. Several investigations and assessments, such as physical and neuro-biological exams, the Mini-Mental State Examination (MMSE), and the patient's thorough history, are required for a proper diagnosis of Alzheimer's disease (AD). Today, doctors use brain MRI images to diagnose Alzheimer's disease. MRI or Magnetic Resonance Imaging is a technique of medical imaging used to produce images of the physiological processes of the body. The data available in the form of MRI can be processed to extract the specific required data for the purpose of the experiment by using image processing techniques and this extracted data can then be used in order to analyze and generate the results.

Throughout the years, there has been various studies and research surrounding Alzheimer's Disease. Various applications in the field of biomedical imaging have been developed as a result of the same. Tools for brain image analysis, brain extraction and a lot more are currently available to us due to the advancement in the field of biotechnology and related research [4], [5], [6], [7], [8].

In [9], Rajesh et. al. presents a strategy to detect and extract brain tumour from patient's MRI scan images of the brain. A fully automatic segmentation of the brain in MRI is represented in [10]. In [11], [12], [13] reviewed various brain MRI segmentation methods covering various modalities, methods for noise reduction during the process, inhomogeneity correction and segmentation. This review also mentions MRI segmentation as a challenging task and highlights a need for research to improve accuracy, precision, and speed of segmentation methods. It also suggests the use of improved atlas-based[14] methods for bringing improvements in the brain segmentation methods. Improvement in the accuracy of the segmentations can be achieved by selecting atlases from large databases is mentioned in [15].

At the same time, machine learning algorithms and models are also applied in AD classification[16]. Zhu et al. shows a method for feature selection and classification based on SVM in [17], [18]. A combination of Principal component Analysis (PCA) for feature selection and Support vector machine Radial basis function (SVM-RBF) for classification is presented in [19], [18]. Deep learning for

E-mail address: ashalatha.20phd7020@vitap.ac.in, namburi.anupama@vitap.ac.in

the purpose of AD classification has also been researched upon. A deep Neural Network was designed to predict mild cognitive impairment (MCI) conversion to AD using MRI data [20], [21]. An approach based on Stacked Auto-Encoder(SAE) along with a multi-kernel SVM was also studied and is presented in [22].In [23] hybrid Convolution neural network (CNN), [24] generative adversarial learning, Fine tuned residual neural network (ResNet) [25] are proposed.

In this paper, the data is configured and process per the requirement and then implemented deep network classification techniques and also compare it with the existing techniques like K-NN, decision tree, SGD and SVC classifiers to identify the differences in each of these methods. Tools and Various standards used in this paper are:

- Neuroimaging Informatics Technology Initiative (NIfTI) is basically an open file format used to store brain image data using MR Imaging technology.
- FMRIB Software Library(FSL) has Brain Extraction Tool(BET) to detecting and deleting non-brain tissue from an input image [26], [27], [28].
- Statistical Parametric Mapping(SPM 12) is a voxel-based approach, to identify specialized regions in the brain and is the most common approach for characterizing functional anatomy and disease related changes [29].

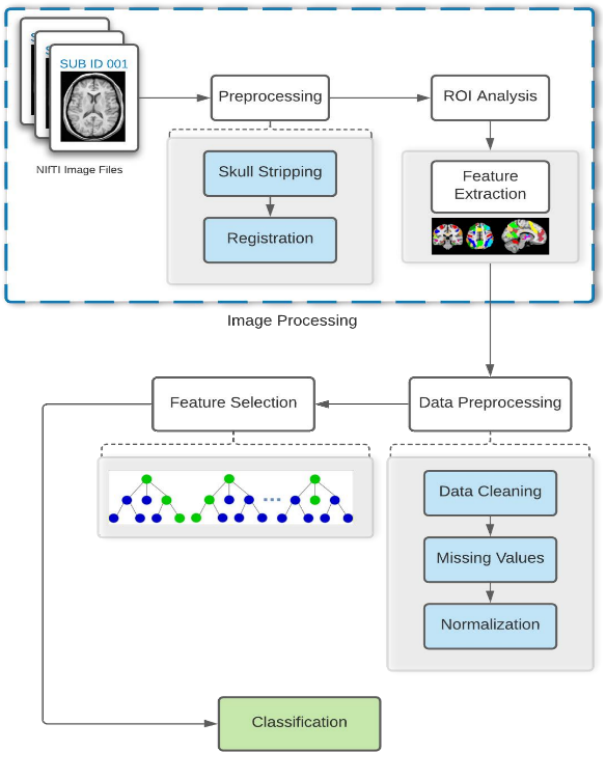


Figure 1. AD ALERT block diagram

2. Research Methodology

The proposed system is divided into three major sections. Starting with the image acquisition and processing section followed by the data processing section and finally the feature selection and classification section.

The first section is the image processing section. It consists of three sub sections – Image Acquisition, Pre-processing and ROI Analysis. The second section presents the extracted raw data processing, including – data cleaning, missing value analysis and normalization of data. In the final section, we perform comparative analysis of various models for feature selection (importance determination) and for the classification model, an analysis on both traditional and deep network models are presented. Each of these is elaborated in the upcoming sections. The block diagram of the proposed methodology is shown in Fig. 1.

A. Image Data Acquisition

Data used in the preparation of this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuro psychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Table I shows the image data acquisition statistics

TABLE I. Image data acquisition statistic

Class Label	Gender	Age Range	Percentage
CN	Male	79 to 89	38%
	Female	71 to 94	62%
AD	Male	68 to 86	48.20%
	Female	71 to 87	51.80%

– the gender-based age ranges and percentage of subjects available. About 357 CN (Cognitive Normal) and 212 AD (Alzheimer's Dementia) MRI scans were taken into consideration for the analysis of this paper.

B. Pre-processing

1) Skull Stripping

Skull stripping is the process of removing non-brain tissues from the image files that are available and is a very vital step in the pre-processing section as it enables accurate measurements and mapping in the further processes. Skull stripping was performed with the help of FSL BET tool and the steps for the same are as follows –

- BET can be used both directly through the command line or by using the GUI option. For the purpose of illustration, it is presented in this paper using the GUI..
- The GUI application is accessed by entering the command fsl in the terminal.

- Next step is to perform the brain extraction. For the sole purpose of demonstration, the brain extraction on a single image (NIfTI file) is performed. The BET uses brain extraction option from the available list, which in turn opens a dialog box to input the image.
- Select the NIfTI Image (.nii) file and it automatically renames and sets a path for the output file.
- Once everything is set up and good to go, click on "Go". The command line automatically executes a command for bet and the process is complete . A new file is created with the same file name as of the input with a postfix "brain" added to it (eg: test4.nii – input file; test4-brain.nii.gz – output file).
- Below Fig.2, shows the original input image and Fig.3 shows the extracted brain image (highlighted in blue).

Figure 2. Original Image (Sample)

Figure 3. Extracted Output Image (Sample)

Finally, from this process, extracted brain is available for output or simply the process of skull stripping is complete. The above steps show the extraction for a single input, but in the case of this experiment the number of available inputs were large and hence they were executed directly via the command line in batches.

2) Registration

Registration [4] is the process used in mapping the subject's data to a common coordinate system in which anatomical features are in alignment. This process thus helps in comparing anatomical and brain image data across multiple subjects all in the common space, thus making the further region-specific analysis consistent and accurate. Talairach normalization was one of the very first commonly used volumetric alignment techniques [4]. For the purpose of registration, a toolbox called WFU PickAtlas in combination with SPM 12 is used in MATLAB, which in fact provides a method for ROI masks generation based on the same Talairach Daemon database. The process is combined with Region of Interest (ROI) Analysis and is described in the following section.

C. ROI (Region of Interest) Analysis

ROI Analysis as mentioned in [5] is performed and finally the voxels are extracted. For the purpose of ROI analysis and feature extraction, SPM 12 was followed using MATLAB. The following steps are followed to work with SPM after the installation.

- In the MATLAB Command Window, enter the command spm and execute it.
- On successful execution of the command, we have an application open up. This is the SPM 12 GUI that can be used to perform various analysis, segmentation processes, registration etc.
- Fig. 4 shows a view of an anonymous subjects' skull stripped brain ready for analysis.
- The graphic module of this software can be used to view various slices of the brain as well, as seen in Fig.4
- WFU PickAtlas is used to extract ROI masks and the first step in the process is of selecting the appropriate atlas. Once this step is complete, access various regions of the brain as options and generate region specific masks as shown in Fig. 5.
- In Fig. 5, as an example, the brain area – Hippocampus was selected and was moved to the working region.
- Now, this region is highlighted (in red) in the atlas . Finally, save the mask in the specified location.
- Fig .6 is a view of the results from of example ROI Analysis. The values are masked due to confidentiality issues.

This mask is then used to extract the voxel values for each subject considered and the corresponding values were extracted into a .csv file, which is further used as input for the process of classification. All subjects were processed in batches through the command line due to the high number of image files.

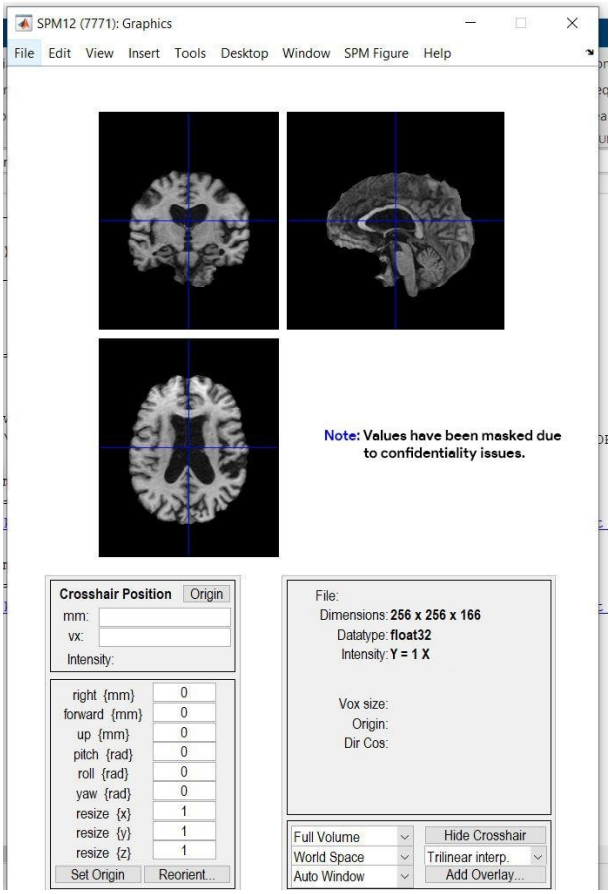


Figure 4. Skull Stripped Subject – All Views

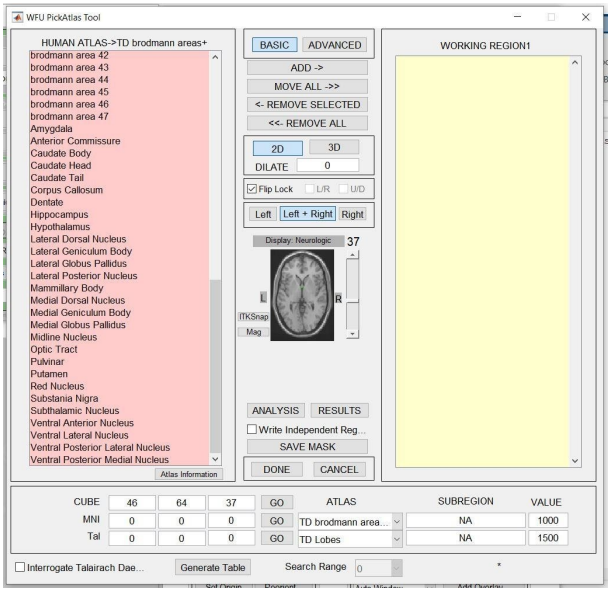


Figure 5. WFU PickAtlas – Tool View

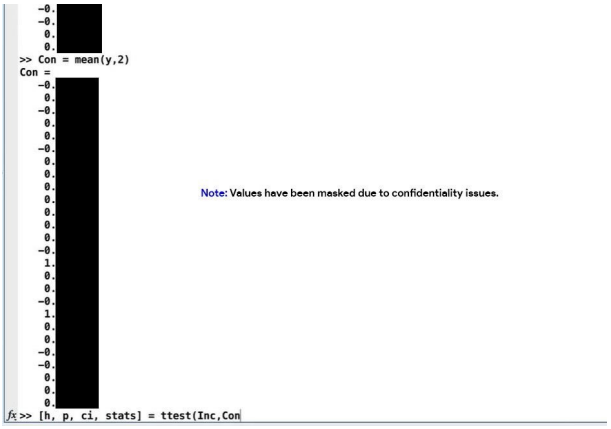


Figure 6. ROI Analysis Result View

D. Data Pre-processing (Preparation)

Now that the image processing phase is completed and have the raw numeric data in hand, further data pre-processing section is to be performed. This section is divided into the following parts – data cleaning, missing value analysis and normalization.

1) Data Cleaning

A basic but extremely essential part of data pre-processing is data cleaning. This process involves detection and removal of errors and other inconsistencies in the extracted data. For the same purpose, missing value analysis is performed, filling of the missing value and a normalization process.

TABLE II. Extracted Raw Data-Sample Schema.

Id	Label	f1	f2	f3	f4	.....	fn
1	CN	18.3	0.15	0.8	0.23	.....	-
2	AD	0.1	0.8	NaN	0.65	.....	-
3	AD	0.12	NaN	0.23	0.76	.....	-
4	CN	20.4	0.16	0.8	0.2	.....	-
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
n	CN	18.29	0.11	NaN	0.3	.....	-

2) Missing Values

In the data extracted, some features are missing at very specific points. Consider the Table II that shows these specific points, for example – row 2, column f3 – the value corresponds to NaN.

Points like these need to be replaced with numeric data. Two techniques in this case are applicable– either fill in all these missing points with zeros '0' or replace with the average value of the feature for the same class label. In the case of missing values of the extracted data, the former technique is considered. After replacing the missing values in the data, the data set look like as shown in Table III.

TABLE III. Detecting and replacing the missing values of table II.

Id	Label	f1	f2	f3	f4	.....	fn
1	CN	18.3	0.15	0.8	0.23	.....	-
2	AD	0.1	0.8	0	0.65	.....	-
3	AD	0.12	0	0.23	0.76	.....	-
4	CN	20.4	0.16	0.8	0.2	.....	-
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
n	CN	18.29	0.11	0	0.3	.....	-

TABLE IV. Normalization of table III.

Id	Label	f1	f2	f3	f4	.....	fn
1	CN	0.8965	0.15	0.8	0.23	.....	-
2	AD	0	0.8	0	0.65	.....	-
3	AD	0.0009	0	0.23	0.76	.....	-
4	CN	1	0.16	0.8	0.2	.....	-
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
n	CN	0.8961	0.11	0	0.3	.....	-

3) Normalization

From the Table III it can be observed that different features represented in different scales. Feature 'f1' has values ranging between 0 and 21, while feature 'f2' has values ranging between 0 and 1. The process of adjusting values such that the entire data set falls under a common scale is called as normalization. The most commonly used normalization technique is the min-max technique, and it is best in cases where the boundary values maximum and minimum are known [6] and the formula for the same is given as follows

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

x is the feature value, x_{min}, x_{max} are the minimum and the maximum values of the features. An example of the normalization process is shown in Table IV. All the features in the table are now on a common scale varying between 0 and 1.

E. Feature Importance and Selection

Feature selection helps to select a limited set of variables from the available set of features that can efficiently describe the input data. It also helps in dimensionality reduction. The other benefits of feature selection include understanding the data better, reduction in computation requirement, improvement in the predictor performance, et cetera.

Before moving into the process, the available set of features and value boundaries were observed. Fig. 7 and 8,present the features – f1 to f10, f11 to f20, f21 to f30, respectively. The variance of the data points can be seen in these box plots, label wise. The box plot shown in Fig. 7 depict that the values of the features f1 to f9 related to the Alzheimer cases are higher than the values of the features f1 to f9 of normal cases except f10 feature.

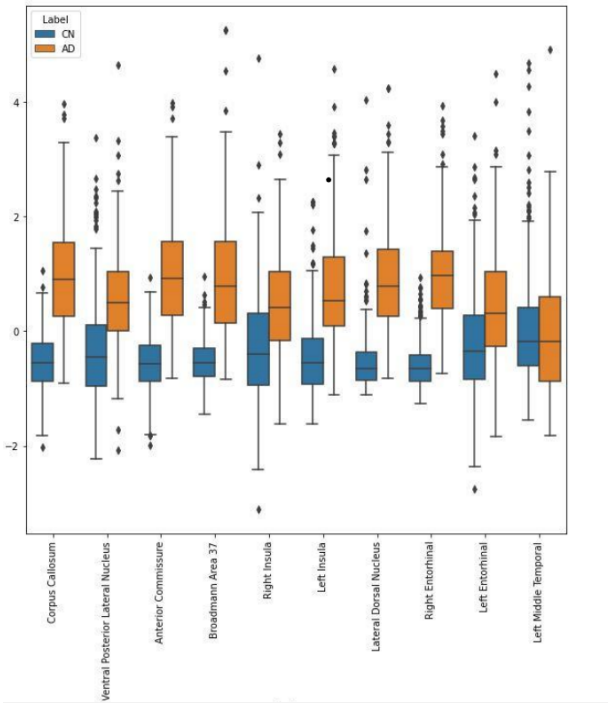


Figure 7. Feature values (f1-f10) - Box Plot

Similarly, the box plot shown in Fig.8 a infers that the all the features of Alzheimer are greater than normal cases except f5 and f9. The features f21 to f30 has all the features values of Alzheimer cases are greater than normal cases are seen in Fig.8 b. From these figures it can be concluded that the feature values of Alzheimer cases on an average are greater than the feature values of the normal cases.

To observe the correlation between the features, a heat map is plotted as shown in Fig.9. The darker shade indicates the lesser correlation, and the lighter color shows the highest correlation. This gives us a better understanding of the correlated coefficients between variables, for example the f1 (broadmann area 41) has very less correlation with f7(broadmann area) 32,f17(broadmann area 4),and f25(broadmann area 28) and has better correlation with other features.

To determine the important and find the optimal features to define the dataset, feature importance and selection technique using Random Forest is performed. The reason for considering random forest rather than decision tree is because in case of the later, a decision tree is built on entire dataset with all the available features, but in case of the former, it randomly selects instances in the dataset to build various number of trees and finally generate the results based on the average and voting mechanism.

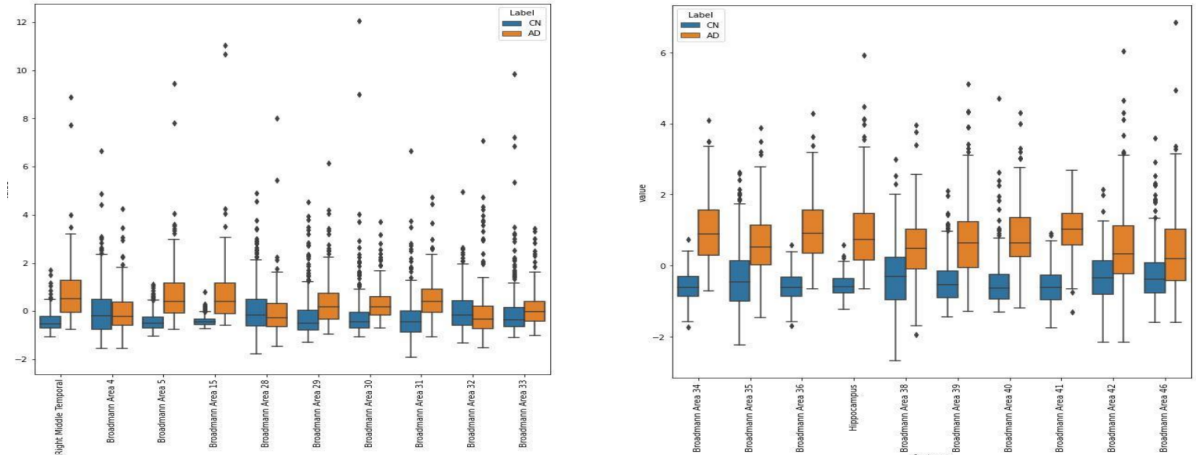


Figure 8. (a). Feature values (f11-f20) - Box Plot. (b) Feature values (f21-f30) - Box Plot

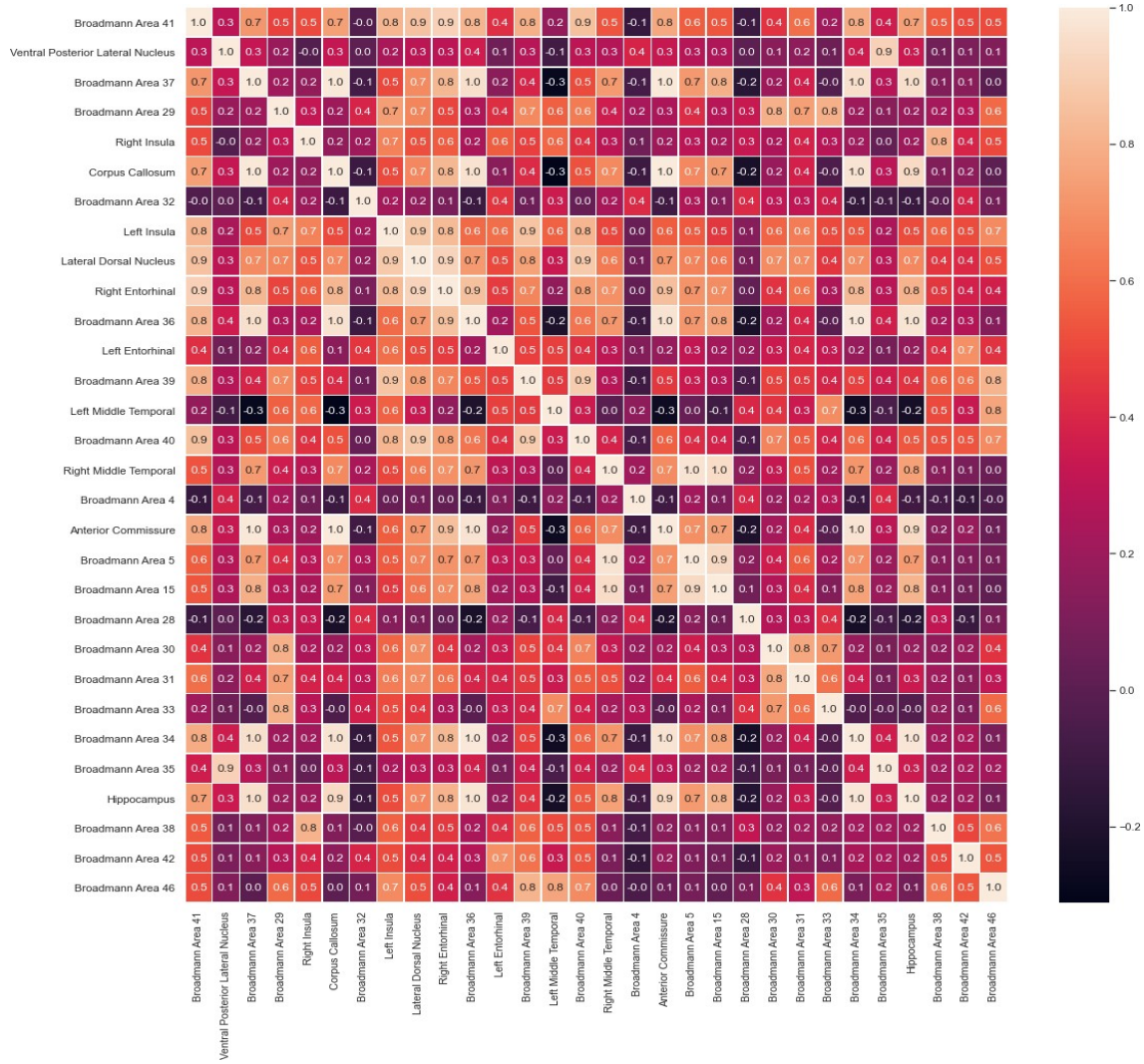


Figure 9. Correlation Map

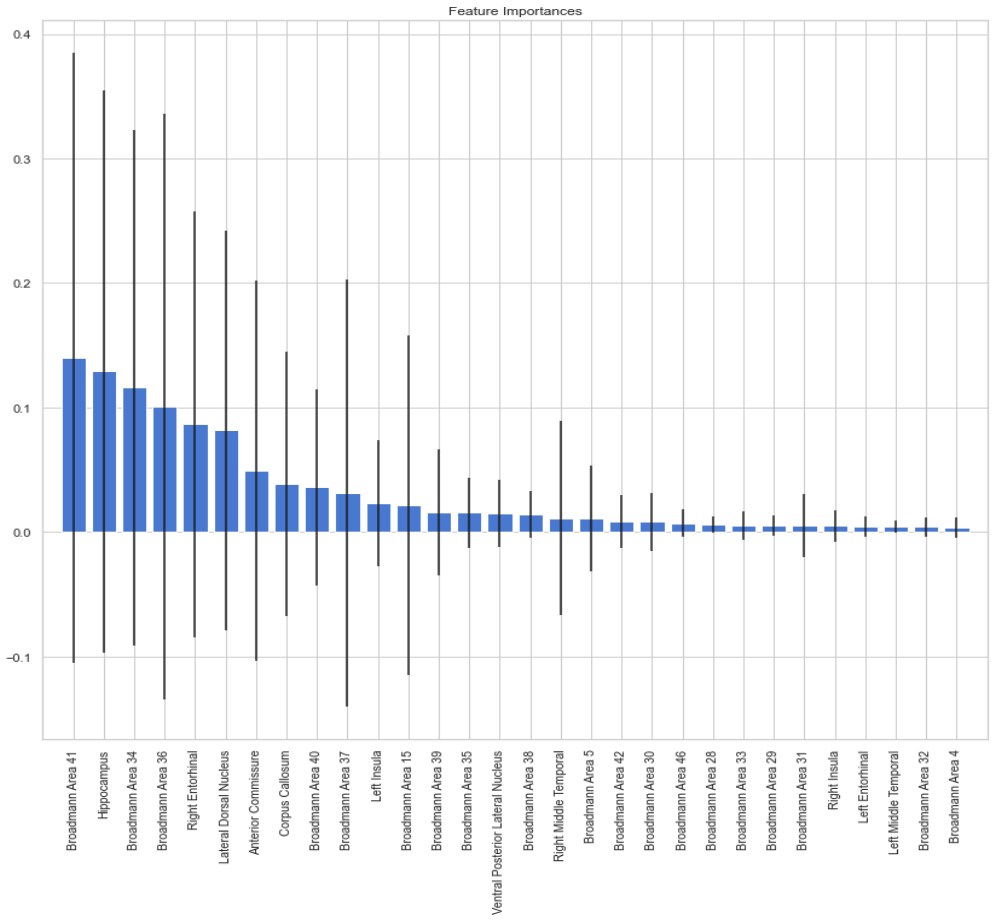


Figure 10. Feature Importance - Results

Using this technique, the forests try and pick nodes that have high fraction of instances from the same class label by finding features that clearly divide the dataset into the available set of classes. The node split mechanism is based on Gini Impurity [8] and is given by

$$G = \sum_{i=1}^{C} p(i) * (1 - p(i)) \tag{2}$$

where c represents the class count and $p(i)$ indicates the ratio of class at i^{th} node.

So, in this case even if one tree is susceptible or sensitive to noise, since multiple trees exist the tree correlation and noise sensitivity decreases when the result is an average of all of these trees [8]. This makes the random forest technique better than the decision tree technique and other methods when it comes to feature importance and selection.

On implementation of the same, as a result a table of feature importance is generated ranking the most important feature the highest value and the least important feature the lowest value and the table looks similar to Table V. The results from the feature importance table were plotted on a

TABLE V. Feature Importance(Sample Result Table)

	Feature Name	Feature Importance
24	f25	0.695
0	f1	0.145
25	f26	0.09
26	f27	0.011
9	f10	0.009
14	f15	0.0070
27	f28	0.0073
6	f7	0.00734
4	f5	0.00702
12	f13	0.0069
1	f2	0.0064
18	f19	0.00564
19	f20	0.00205
.	.	.
.	.	.
.	.	.
28	f29	0

graph to visualize the result and is shown in Fig. 10. The top ranked 30 percentage and optimum features out of the total extracted features are considered for the classification.

3. Classification

To classification of the data, a comparative analysis on both traditional and deep network models were performed. This section will give a deeper insight into the methods studied and applied.

A. Traditional Classification

Multiple traditional classification models have been studied and tested on the data during this paper timeline and the observations were analysed and projected in the results section of this paper.

The following models were used for the same –

1) K-NN Classifier.
2) Decision Tree Classifier.
3) SGD Classifier
4) SVC

1) K-NN Classifier

The Nearest Neighbour Classifier identifies the nearest neighbours to a query example and uses those neighbours to determine the class of the query or the newly introduced data point with the help of a voting mechanism. Here k nearest neighbours is considered to determine the class of the new data point.

On obtaining the query's k-nearest neighbour's distance, we move onto the voting mechanism and there are majorly two common approaches that are followed for the same and are-

Inverse distance – weighted voting, where the closer the neighbours the higher the votes they get. Majority voting, all votes are considered equal here. For each class c ϵ L, we count how many of the k neighbours y_j have that y_c class. The class with the most votes is returned.

Another approach to voting is based on Shepard's work and uses an exponential function rather than inverse distance, i.e:

$$vote(y_j) = \sum_{c=1}^{k} e^{-d(q,x_c)} 1(y_j, y_c) \quad (3)$$

As a result, neighbour x_c vote for class y_j is equal to one divided by the distance between the neighbour and the unknown example q, and $1(y_j, y_c)$ returns one if the class labels match and zero otherwise.

2) Decision Tree Classifier

Because of the training speed and transparent model production, decision tree algorithms are pretty popular in the field of data mining [29], [30].

As the name suggests, decision tree classifiers are tree based that classify the labeled trained data into a tree or rules. On derivation of the tree in the model learning phase, random instances are selected for accuracy testing after which unlabeled instances from the testing data are passed through the tree model to classify them into either of the labels available.

In the proposed methodology, Gini Impurity index is used for split criterion. Gini Index is a cost function used to evaluate splits in the dataset and given by the equation 2.

3) SGD Classifier

A linear classifier optimized with the help of Stochastic Gradient Descent. With a decreasing learning rate, gradient of the loss is estimated each instance at a time and the model is updated along the way.

Scikit-learn implementation of the same for the purpose of SGD classifier.

4) Support Vector Classifier (SVC)

SVC, somewhat like SGD works as an optimizer and tends to draw a straight line between two labels such that the distance between these two are as far as possible. The aim here is to maximize the width.

The scikit-learn implementation of the same for the purpose of our experiment. The parameters that are used for the same as follows- 'C':0.5, 'gamma':0.01 Where 'C' is a hyper parameter in SVM to control error and Gamma- decides the curvature of the decision boundary. the values for the regularizing parameters C=0.5 and gamma=0.01 worked best in for the classification based on trial-and-error method.

B. Deep Network(DN) Models

Deep Neural networks are said to have achieved amazing performance in the field of machine learning. Deep Neural networks (DNNs) is comprised of input layer, an output layer, and numerous hidden layers. Each node in a layer is connected to every other node in the previous layer using a predefined set of weights. Weight selection for training the network is crucial for each layer. Values for each layer are computed from the values of the preceding layer, resulting in values for the outputs in the final, by assigning values to the inputs and then feeding them forward through the network. In this work based on the type of data and class labels, DNN's under various settings and architectures were tried out and the details are mentioned in the next sections.

DNN's have an advantage that it makes complex natural systems easy to handle and model them appropriately. In DNN's non-linear activation function is used after calculating a linear combination of the node values from the preceding layer to determine the value of each hidden node in the network. Rectified Linear Unit (ReLU) is used as activation function. A node's value is calculated when a ReLU activation function is applied to it as the maximum of the linear combination of its nodes from the previous

layer and 0. As a result, we can think of ReLUs as the function

$$ReLU(x) = max(0, x) \quad (4)$$

Consider input layer is layer1, the output layer is layer n, and the hidden layers are levels $2, ..., n-1$. Let X_{ij} be the value of j^{th} node in i^{th} layer, W_i be the weight matrix at i^{th} layer and b_i.For $2i < n$, X_{ij} is given by$X_{ij} = ReLU(W_iV_i1 + b_i)$, with the ReLU function being applied element-wise. The output layer is calculated by the sigmoid transfer function which is given by

$$sg(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

The architecture proposed in this work has two hidden layers having ReLU as the transfer function and the output layer having a transfer function of Sigmoid. Four models are created in the proposed work. Model A has an architecture of 16-16-20-1 nodes input, 2 hidden and output layer. Model B is constructed with the same architecture with batch normalization and adding a dropout of 0.3. Model C is constructed with an architecture of 16-26-20-1 nodes input, 2 hidden and output layer. Model D is constructed with the same architecture with batch normalization and having dropout of 0.3. Here 0.3 is the maximum dropout allowed in the deep networks. The architectures of Model A and Model B are shown in figures 11 and 12.

Model: "sequential_1"

Layer (type)	Output Shape	Param #
dense_3 (Dense)	(None, 16)	272
dense_4 (Dense)	(None, 20)	340
dense_5 (Dense)	(None, 1)	21

Figure 11. Model A

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 16)	272
dropout (Dropout)	(None, 16)	0
batch_normalization (BatchNo	(None, 16)	64
dense_1 (Dense)	(None, 20)	340
dropout_1 (Dropout)	(None, 20)	0
batch_normalization_1 (Batch	(None, 20)	80
dense_2 (Dense)	(None, 1)	21

Figure 12. Model B

4. Results and Discussions

A comparative study between both traditional and deep network models were performed. The performance metrics are defined based on the following terms.

- True Positive (TP) is when the prediction is positive, and it is true.
- True Negative (TN) is when the prediction is negative, and it is true.
- False Positive (FP) is when the prediction is positive, but it is false.
- False Negative (FN) is when the prediction is negative, and it is false.

The metrics used for the purpose of generating results and classification report are as follows (i) Accuracy

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \quad (6)$$

(ii) Precision – A measure of the rate of true positives among all detections.

$$Precision = (TP)/(TP + FP) \quad (7)$$

(iii) Recall – Measures the percentage of detected ground truth annotations.

$$Recall = (TP)/(TP + FN) \quad (8)$$

(iv) F1-score – A weighted harmonic mean of precision and recall used to measure the percentage of correct positive predictions.

$$F1 = (2RecallxPrecision)/(Recall + Precision) \quad (9)$$

(v) ROC-AUC ROC is the probability curve and Area under the curves is the degree or a measure of separability. The higher the ROC-AUC curve the better the classification accuracy.

TABLE VI. Confusion matrix of traditional models

		Normal	Alzheimer	Total
K-NN	Normal	108	0	108
	Alzheimer	5	102	107
Decision Tree		Normal	Alzheimer	Total
	Normal	107	1	108
	Alzheimer	8	99	107
SGD		Normal	Alzheimer	Total
	Normal	107	1	108
	Alzheimer	3	104	107
SVC		Normal	Alzheimer	Total
	Normal	107	1	108
	Alzheimer	4	103	107

The confusion matrix of the traditional models are shown in VI.The tradition models namely the K-NN, decision tree, SGD and SVC are experimented on the features selected using the random forest. The study considered 108 normal cases data and 107 Alzheimer data.

The K-NN correctly classified the normal cases but misclassified 5 Alzheimer cases as normal cases resulting in an accuracy if 0.9767. The decision tree classifier wrongly classified 1 normal case as Alzheimer case and 8 Alzheimer cases as normal cases. The SGD classifier wrongly classified 1 normal case as Alzheimer case and 3 Alzheimer case

TABLE VII. Analysis of various Deep Network Architectures

Deep Network Model	Transfer Function	Architecture	Dropout	Batch Normalization	Accuracy
A	ReLU	16-16-20-1	0	-	0.9649
B	ReLU	16-16-20-1	0.3	YES	0.9649
C	ReLU	16-26-20-1	0.3	YES	0.9825
D	ReLU	16-26-20-1	0	-	0.9737

TABLE VIII. Model Accuracy comparison with other existing models

Model	Accuracy
K-NN	0.9767
K-NN[31]	0.9515
Decision Tree	0.9581
Decision Tree[32]	0.77
SGD	0.9814
SVC	0.9720
DN Models	0.9825

TABLE IX. Performance metrics

Model	**Precision**	**Recall**	**F1 score**	**ROC-AUC**
K-NN	1	0.9532	0.976	0.9766
SGD	0.9904	0.9719	0.981	0.9813
DT	0.99	0.9252	0.9565	0.9579
SVC	0.9904	0.9626	0.9763	0.9766
DN Model	0.99	0.9724	0.982	0.9824

as normal cases. The SVC classifier wrongly classified 1 normal case as Alzheimer case and 4 Alzheimer cases as normal cases.

The four different models of deep neural networks created are experimented on the features selected with the random forest. The four models and their architecture, batch normalization and the dropout considerations are presented in Table VII. It shows the with the selected configurations, DN Models C and D give the highest accuracy of 0.9825 and 0.9737. With the batch normalization and the dropout added Model C shows better result than the other deep models considered.

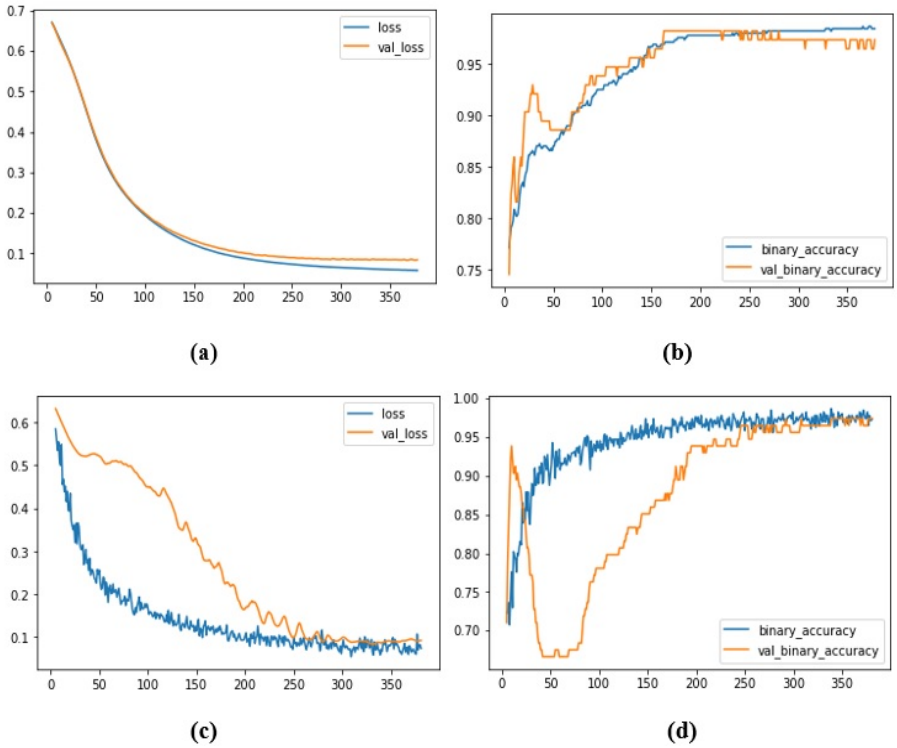


Figure 13. Corresponding graphs for Models A and B in Table VI - (a) Loss and validation loss for Model A, (b) Accuracy graph for Model A, (c) Loss and validation loss for Model B, (d) Accuracy graph for Model B

The comparison of the traditional models and the deep network model is given in Table VIII. From the comparison the deep learning model C achieves good accuracy when compared to traditional models. Also, the feature selection based on the random forest has improved the accuracies of the tradition model with accuracy K-NN achieving 0.9767, decision tree of 0.9581, SGD of 0.9814 and SVC of 0.9720.

A study with various feature extraction methods based on K-NN was analyses in paper [31]. The author has analysed 7 different transform domain method for feature selection among which dual tree complex wavelet-based feature selection achieved height with 0.9515 accuracy. With random forest selection method K-NN achieved 0.9767 accuracy. The improved decision tree algorithm proposed in [32] achieved an accuracy of 0.77 which is less when compared to the result obtained with decision tree with feature selection. From these comparisons it can be infer that the feature selection based on the random forest provides better results in classification.

Table IX shows the different performance metrics namely precision, recall,f1Score and the ROC_AUC . Even though the precision for the K-NN model is high but the recall, f1score and the ROC_AUC values are less. The decision tree shows the less performance with respect to the considered parameters. The SGD from the traditional model and the DN models shows better results compared to the other models considered. The DN models outperforms all other models with respect to recall, f1score and the ROC_AUC values.

The figure 13 shows the plots (a) Loss and validation loss for Model A, (b) Accuracy graph for Model A, (c) Loss and validation loss for Model B, (d) Accuracy graph for Model B. From the plots (a), (c) it can be observed that the train and the testing loss is gradually reduced indicating that the model is not over fitted or under fitted. Also, from the graphs (b) and (d) it can be observed the both the training and the testing accuracies are very good and not varying much indicating the consistency of the model. Considering all the performance metrics, it is clearly observed that the AD Alert based on DN models with feature selection using random forest is suitable for Alzheimer disease classification, making it affordable and available for efficient prediction as it analyze MRI image data itself. Using the AD Alert the practitioner can easily predict the AD disease and make decisions about the pathology in less time.

5. Conclusions and Future Work

One of the most severe neurological conditions is Alzheimer's disease. Researchers are working to develop a precise and affordable method for classifying AD using brain imaging because the conventional diagnosis process is costly and occasionally erroneous. AD Alert is one step towards the precise and affordable method to predict the AD at low cost. The paper has presented different AD classification algorithms that are implemented with the features selection based on random forest. The traditional classification algorithms have improved their performance with use of random forest feature selection. The paper also presented the DN models for AD classification with performance comparisons and observations. The comparison of the findings reveals that the classification method based on DNNs can yield the most convincing results (approximately 98.25%). The paper used binary classification that can be extended to multi class classification problem of AD with the use of random forest feature selection and can achieved better performance.

References

[1] [Online]. Available: https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/

[2] M. Monica Moore, M. Díaz-Santos, and K. Vossel, "Alzheimer's association 2021 facts and figures report," *E-Newsletter, Mary S. Easton Center for Alzheimer's Disease Research at UCLA*, 2021.

[3] A. Dement, "Alzheimer's disease facts and figures," *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, vol. 12, no. 4, pp. 459–509, 2016.

[4] S. Smith, P. R. Bannister, C. Beckmann, M. Brady, S. Clare, D. Flitney, P. Hansen, M. Jenkinson, D. Leibovici, B. Ripley *et al.*, "Fsl: New tools for functional and structural brain image analysis," *NeuroImage*, vol. 13, no. 6, p. 249, 2001.

[5] J. Ashburner, G. Barnes, C.-C. Chen, J. Daunizeau, G. Flandin, K. Friston, S. Kiebel, J. Kilner, V. Litvak, R. Moran *et al.*, "Spm12 manual," *Wellcome Trust Centre for Neuroimaging, London, UK*, vol. 2464, 2014.

[6] E. S. Lutkenhoff, M. Rosenberg, J. Chiang, K. Zhang, J. D. Pickard, A. M. Owen, and M. M. Monti, "Optimized brain extraction for pathological brains (optibet)," *PloS one*, vol. 9, no. 12, p. e115551, 2014.

[7] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging," *IEEE transactions on medical imaging*, vol. 36, no. 11, pp. 2319–2330, 2017.

[8] X. Hong, R. Lin, C. Yang, C. Cai, and K. Clawson, "Adpm: An alzheimer's disease prediction model for time series neuroimage analysis," *IEEE Access*, vol. 8, pp. 62 601–62 609, 2020.

[9] R. C. Patil and A. Bhalchandra, "Brain tumour extraction from mri images using matlab," *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE)*, vol. 2, no. 1, p. 1, 2012.

[10] M. S. Atkins and B. T. Mackiewich, "Fully automatic segmentation of the brain in mri," *IEEE transactions on medical imaging*, vol. 17, no. 1, pp. 98–107, 1998.

[11] I. Ahmed, Q. N. U. Rehman, G. Masood, M. Nawaz, A. Adnan, and M. Khan, "Segmentation of brain tumor from healthy tissues using multimodal mri images," *International Journal of Computer Science and Information Security*, vol. 14, no. 10, p. 676, 2016.

[12] S. Harish, G. A. Ahammed, and R. Banu, "An extensive research survey on brain mri enhancement, segmentation and classification," in *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*. IEEE, 2017, pp. 1–8.

[13] K. Ramesh, G. K. Kumar, K. Swapna, D. Datta, and S. S. Rajest, "A review of medical image segmentation algorithms," *EAI Endorsed Transactions on Pervasive Health and Technology*, 2021.

[14] E. Buser, E. Hart, and B. Huenemann, "Comparison of atlas-based and neural-network-based semantic segmentation for dense mri images," *arXiv preprint arXiv:2109.14116*, 2021.

[15] L. Sun, L. Zhang, and D. Zhang, "Multi-atlas based methods in brain mr image segmentation," *Chinese Medical Sciences Journal*, vol. 34, no. 2, pp. 110–119, 2019.

[16] S. Mahajan, G. Bangar, and N. Kulkarni, "Machine learning algorithms for classification of various stages of alzheimer's disease: A review," *Machine Learning*, vol. 7, no. 08, 2020.

[17] T. Dening and M. B. Sandilyan, "Dementia: definitions and types," *Nursing Standard (2014+)*, vol. 29, no. 37, p. 37, 2015.

[18] M. Sudharsan and G. Thailambal, "Alzheimer's disease prediction using machine learning techniques and principal component analysis (pca)," *Materials Today: Proceedings*, 2021.

[19] R. N. Kalaria, G. E. Maestre, R. Arizaga, R. P. Friedland, D. Galasko, K. Hall, J. A. Luchsinger, A. Ogunniyi, E. K. Perry, F. Potocnik *et al.*, "Alzheimer's disease and vascular dementia in developing countries: prevalence, management, and risk factors," *The Lancet Neurology*, vol. 7, no. 9, pp. 812–826, 2008.

[20] K. Yang, E. A. Mohammed, and B. H. Far, "Detection of alzheimer's disease using graph-regularized convolutional neural network based on structural similarity learning of brain magnetic resonance images," *arXiv preprint arXiv:2102.13517*, 2021.

[21] H. S. Suresha and S. S. Parthasarathy, "Detection of alzheimer's disease using grey wolf optimization based clustering algorithm and deep neural network from magnetic resonance images," *Distributed and Parallel Databases*, pp. 1–29, 2021.

[22] F. Forette and F. Boller, "Hypertension and the risk of dementia in the elderly," *The American journal of medicine*, vol. 90, no. 3, pp. S14–S19, 1991.

[23] Y. Eroglu, M. Yildirim, and A. Cinar, "mrmr-based hybrid convolutional neural network model for classification of alzheimer's disease on brain magnetic resonance images," *International Journal of Imaging Systems and Technology*, 2021.

[24] X. Zhou, S. Qiu, P. S. Joshi, C. Xue, R. J. Killiany, A. Z. Mian, S. P. Chin, R. Au, and V. B. Kolachalama, "Enhancing magnetic resonance imaging-driven alzheimer's disease classification performance using generative adversarial learning," *Alzheimer's research & therapy*, vol. 13, no. 1, pp. 1–11, 2021.

[25] M. Odusami, R. Maskeliūnas, R. Damaševičius, and T. Krilavičius, "Analysis of features of alzheimer's disease: Detection of early stage from functional brain changes in magnetic resonance images using a finetuned resnet18 network," *Diagnostics*, vol. 11, no. 6, p. 1071, 2021.

[26] L. Rizzi, I. Rosset, and M. Roriz-Cruz, "Global epidemiology of dementia: Alzheimer's and vascular types," *BioMed research international*, vol. 2014, 2014.

[27] P. Cunningham and S. J. Delany, "k-nearest neighbour classifiers: (with python examples)," *arXiv preprint arXiv:2004.04523*, 2020.

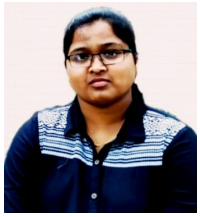
[28] G. H. Wong and D. V. Goeddel, "Induction of manganous superoxide dismutase by tumor necrosis factor: possible protective mechanism," *Science*, vol. 242, no. 4880, pp. 941–944, 1988.

[29] D. Lavanya and K. U. Rani, "Ensemble decision tree classifier for breast cancer data," *International Journal of Information Technology Convergence and Services*, vol. 2, no. 1, pp. 17–24, 2012.

[30] J. Pettersson, "The use of decision trees to detect alzheimer's disease," 2021.

[31] U. R. Acharya, S. L. Fernandes, J. E. WeiKoh, E. J. Ciaccio, M. K. M. Fabell, U. J. Tanik, V. Rajinikanth, and C. H. Yeong, "Automated detection of alzheimer's disease using brain mri images–a study with various feature extraction techniques," *Journal of Medical Systems*, vol. 43, no. 9, pp. 1–14, 2019.

[32] A. Kumar and T. R. Singh, "A new decision tree to solve the puzzle of alzheimer's disease pathogenesis through standard diagnosis scoring system," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 9, no. 1, pp. 107–115, 2017.



Vunnam Asha Latha Vunnam Asha Latha is currently perceiving her PhD under the guidance of Dr.Anupama Namburu in the school of Computer Science and Engineering, VIT-AP University, Amaravathi, Guntur, India. She completed her Masters from the department of Computer Science and Engineering in Acharya Nagarjuna University, Guntur.

Anupama Namburu Dr. Anupama Namburu is working as an associate professor in the school of Computer Science and Engineering, VIT-AP University, Amaravathi, Guntur, India. She worked as an assistant professor in the Department of Computer Science and Engineering in Acharya Nagarjuna University, Guntur. She received her MTech from Andhra University, Visakhaptanam, Andhra Pradesh, India. She completed her PhD in Jawaharlal Nehru Technological University, Kakinada. Her research areas include digital image processing, data science, soft computing techniques and Fuzzy applications. Corresponding author. E-mail: namburianupama@gmail.com