AMultimodal Deep Learning Framework for Prediction of Alzheimer’s Disease

By TANMAY SANTRA

WORD COUNT	5441	TIME SUBMITTED	21-JUL-2022 11:20AM
		PAPER ID	88645865

A Multimodal Deep Learning Framework for Prediction of Alzheimer's Disease

Dissertation submitted in partial fulfilment for the award of the degree of

Master of Technology in Computer Science

by

TANMAY SANTRA

Roll No.: CS2035
M.Tech, 2nd year

Under the supervision of
Prof. Pradipta Maji

Machine Intelligence Unit
Computer and Communication Sciences Devision
INDIAN STATISTICAL INSTITUTE

July 2022

CERTIFICATE

This is to certify that the dissertation entitled **"A Multimodal Deep Learning Framework for Prediction of Alzheimer's Disease "**, submitted by **Tanmay Santra** , having the roll number **CS2035** to Indian Statistical Institute, Kolkata, in partial fulfillment for the award of the degree of Master of Technology in Computer Science is a bonafide record of work carried out by him under my supervision and guidance. The dissertation has fulfilled all the requirements as per the regulations of this institute and, in my opinion, has reached the standard needed for submission.

Date: 21-07-2022

_____

Prof. Pradipta Maji
Machine Intelligence Unit
Indian Statistical Institute, Kolkata

Acknowledgements

First and foremost, I take this opportunity to express my sincere thankfulness and deep regard to *Prof. Pradipta Maji*, for the impeccable guidance, nurturing and constant encouragement that he had provided me during my post-graduate studies. Words seem insufficient to utter my gratitude to him for his supervision in my dissertation work. Working under him was an extremely knowledgeable experience to me. I also thank the CSSC and ISI Library for extending their supports in my different ways in my urgent need.

I shall forever remain indebted to my family, teachers and friends for supporting me at every stage of my life. It is their constant encouragement and support that has helped me throughout my academic career and especially during the research work carried out in the last one year.

Date: 21-07-2022

Tanmay Santra
Roll No.: CS2035
M.Tech, 2nd year
Indian Statistical Institute

Abstract

Alzheimer's disease (AD) is one type of dementia that affects significantly among the populations. In this work, the prediction of different stages of AD is addressed. Early detection of AD is very crucial because patient can take important action to prevent this disease. Three stages of this disease are AD, mild cognitive impairment (MCI), and controls (CN). The MCI is a stage in between AD and CN. Most recent studies use single modality data to make prediction of different stages of AD. The integration of multiple modalities may provide better performance, as compared to single modality, for early detection of AD.

In this work, we use Machine Learning (ML) and deep learning (DL) frameworks to analyze both imaging data (magnetic resonance imaging - MRI) and clinical data to classify patients into three classes, namely, AD, MCI, and CN. Here, we propose a new architecture, which contains convolution layer, pooling layer Autoencoder, multimodal deep framework with cross weight and multilayer perceptron (MLP). The convolution layer and pooling layer are used to extract features from MRI, while Autoencoder is used for clinical data. Then, the features of two modalities are combined using a data integration technique, called multimodal data fusion with cross weights. Finally, MLP is used to classify features of joint representation.

In the current study, we use Alzheimer's disease neuroimaging initiative (ADNI) data set for comparative performance analysis of different models. We show that the proposed deep framework significantly outperforms different mono and multimodal shallow models, in term of accuracy, precision, and recall.

Contents

List of Figures

List of Tables

1 CHAPTER 1: Introduction

In 1907, a German psychiatrist, Alois Alzheimer was the First, who told about the disease. According to his name the disease name became AD. Neurodegeneration is a diseases which affects on brain cells and neurons. Due to loss of brain cells a person who is affected by it forgets different act like cognitive functions, thinking and regular activities.Deficit of Cognitive function can lead to following aphasia, apraxia, and disruption in daily functional activities , that is usually done with the help of the brain. Alzheimer's disease(AD) is a kind of Neurodegeneration disorder. It usually starts in middle and old age. For more details information about AD is discussed below. Accumulation of protein (A,β and Tau) of an AD patient's brain in and around neurons leads to form of plaques of amyloids and neurofbrillary tangles causes a steady deterioration in memory (associated with synaptic dysfunction, brain shrinkage, and cell death in brain). AD is not always related to amyloid plaques but must related to neural loss. The regions of the brain are affected by AD are mainly amygdala ,temporal lobe and hippocampus. In those region changes in the brain very fast as compared to the symptoms that are physically optically discerned in patients. So, it is tough to prevent or diagnose the disease early .Some biomarkers of the brain change at the very beginning, then cognitive deficiency happens. Research about AD tells that brain changing may begin at least 20 years before symptoms appear. The initial stage of AD is called Mild Cognitive Impairment (MCI). MCI is a transitional state in between normal(CN) and AD. However a person having MCI is still able perform regular activities.Two type of MCI patients are MCI converters (MCIc) or MCI non-convertors (MCInc) which means that MCI patients have or have not converted to AD within 18 months. Around 30–40% of MCI patients convert to AD within 5 years.

The main risk factor of AD patients is family background that means AD is related to gene.An important clinical examination is Mini-Mental State Examination(MMSE) score, based on comprehensive interaction of the patient,used to diagnose AD. A survey says that around 0.8 percent of people are affected by AD. Moreover, forty-seven million patients worldwide are affected by it up till now. Thus, diagnosis of is very crucial because early treatments may delay the advancement of AD, which lead that patient can lives more years .
Therefore our main objective is to find different feature extraction techniques which help to diagnosis patient into AD,MCI and CN by development of machine learning, deep learning approaches. Currently the research on computer-aided diagnostics (CAD) has become very popular and intensive. This paper we organize two parts. The first part finds single modality accuracy of MRI and Clinical data. In MRI we use 3DCNN

and for clinical data we apply all classification algorithm like Logistic regression, k nearest neighbor(KNN),Random Forest(RF) ,multilayer perceptron (MLP) etc.Next part we integrate bimodal data by some multimodal technique and also add some statistical feature like GLCM ,first order statistic.

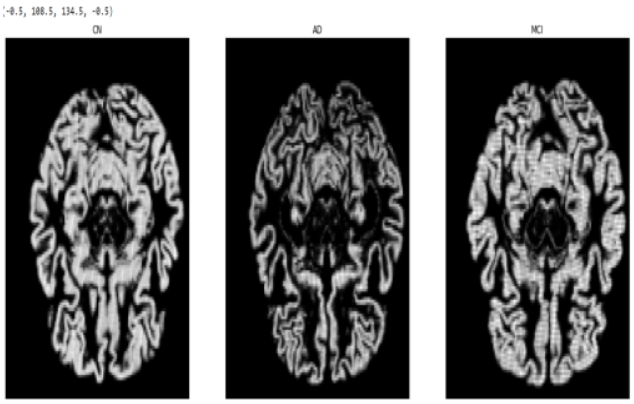

Figure 1: A 2d view of a slice of MRI images of different stages of AD

There are lots of work related to AD among them one of the oldest surveys is classification methods for diagnosis of AD stages by using texture based feature extraction technique on images [2]. In this paper [18] provided a survey of different kinds of feature extraction method and classification method for diagnosis of AD . In the paper [3] have targeted only on feature extraction techniques from given image data. Then authors apply different classifiers for training the model to diagnosis various AD stages. The techniques mentioned by the above authors have become an imperative tool for designing machine learning algorithms to predict neurodegenerative diseases. Multivariate analysis and machine learning techniques for diagnosis of AD used in the paper [4]. This paper also contained a survey of various classification algorithms for diagnosis of the disease. Different types of feature extraction techniques, classification algorithms, and hard challenges for early prediction of AD by focusing on a single modality image dataset (MRI, PET etc) have been demonstrated in the paper [1]. In the paper [12] have presented a review of different kinds of feature extraction method and fusion techniques on those features. In [10], the authors have compared performance of different machine learning algorithms in literature. This paper also demonstrated a computer-aided automated prediction method for the detection of AD stages. In the paper [16], authors have applied different machine learning algorithms on different kinds of neuroimaging datasets.
Feature extraction techniques are important for applying machine learning algorithms that can be applied to diagonosis AD. All papers above the main problem is that authors develop models for predicting stages of AD in comprehensive manner. Some authors have presented dimensionality reduction and classification techniques focusing on different feature extraction algorithms such as texture based, statistical based, etc from neuroimaging data used by classification algorithms to predict disease. Last few years researcher have developed integration of various feature extraction methods and the recent popular trends is related to early prediction of different neurodegenerative disease stages .

There are different technique for feature extraction from MRI data, typically three kind of feature extraction technique are patch- based approach, region of interest (ROI), voxel based approach.
After the feature extraction there is a machine learning approach for classification [8]. Some paper Recently Deep learning has become very popular. Deep learning features are automatically generated without the help of any human expert. Convolution neural network (CNN) based technique used in the paper[7]. Deep learning techniques for medical image processing are discussed in the paper [11] . Multimodal data fu-

sion technique are some paper [14] . Transfer learning for classification used for large dataset. Transfer learning based approach used for classification in the paper [15].

1.1 Motivation

There are lots of models on single modality and very few models developed on multimodal data. So we are interested in multimodal data and mixed different feature extraction techniques like voxel based, biomarker based etc. We will develop new architecture on multimodal data by multimodal deep framework with cross weights.

2 CHAPTER 3: Data Description

Data is very important for experiments. We found data from the ADNI dataset (https://adni.loni.usc.edu/). We collect Structural MRI data and Clinical data from ADNI. In ADNI data repository, There are four kind of studies(ADNI1, ADNI2, ADNI GO, and ADNI3) on imaging, clinical, and genetic data for over 2500 patients . We are focusing on ADNI1, 2 and GO because study of ADNI3 is continuing,expected to end in 2022. The table below gives more details about data [17]

Table 1: Data description

Datatype	Feature	Class label
Clinical data	patient Demography, Different neurological exam, genotype, cognitive task,country origin etc.	AD, MCI, CN
MRI	baseline and cross sectional with T1 weight from ADNI	AD, MCI, CN

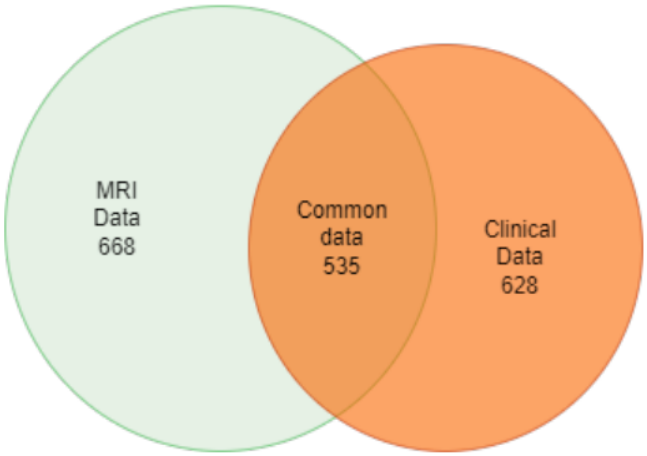

Figure 2: Number of data Found . 668 MRI contains in left side , 628 clinical data contains in right side and 535 patients have common information of both

In this work we use common data type of both Clinical and MRI data for all experiments.

Among 535 patients 108 patients have AD, 268 patients have MCI and 159 patients

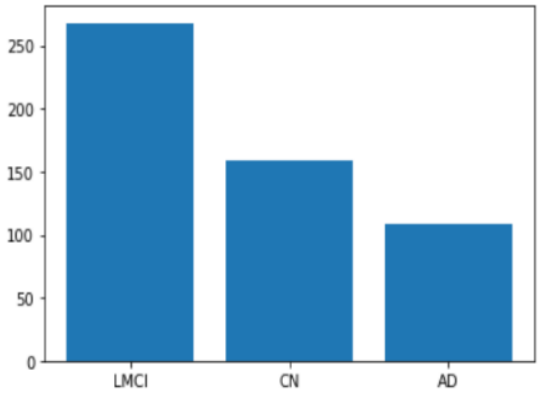

Figure 3: Distribution of AD, CN, MCI

have CN.

Before applying some machine learning technique we need to preprocess the both modality data set. The preprocessing steps we have used are discussed below.

2.1 Data preprocessing

2.1.1 For MRI

- First we reduce noise on MRI data and apply skull striping to get brain tissue by SPM12 software.
- Next We segment brain tissue into Gray matter (GM), white Matter (WM), cerebrospinal fluid (CSF). As most biomarkers tissue like Amygdala, Hippocampus, etc related to AD contain in GM. So we take only GM
- Each Voxel size is different of different GM images. So we normalize GM images into same voxel size and voxel size we give is 1.5mm. Then all GM images normalize to Standard Montreal Neurological Institute (MNI) space.

2.1.2 Clinical data

First we eliminate missing data on clinical data then for categorical data of it converted to numerical value by one hot encoder. Next we standardize Clinical data.

The pictorial view is given below

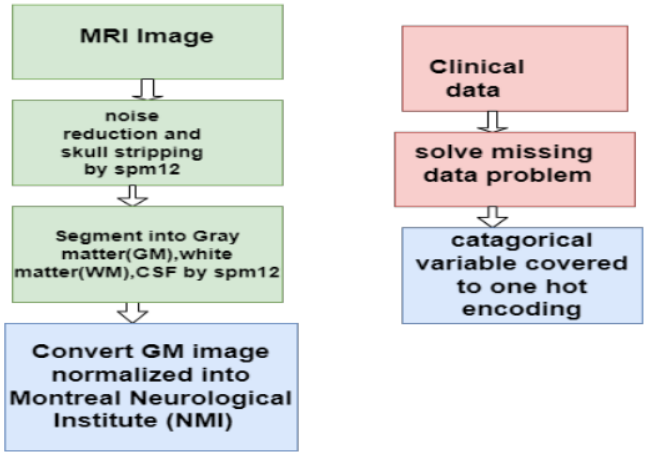

Figure 4: Details of data preprocessing on MRI and Clinical data

Given MRI data have dimension (121,145,121) .Next we select some slices for region of interest(ROI) because some slice contains very small information and some blank space also contains.Finally dimension of a MRI data will be (32,112,96)

3 CHAPTER 4: Method

Here we will show our new architecture for prediction of Alzheimer disease which contains a combination of different methodology like Convolution layers, pooling layers, Autoencoder, Multimodal deep framework for data fusion and finally Multilayer Perceptron(MLP) for classification. We will discuss step by step in the following(and see figure 4)

- After preprocessing, each MRI data has dimension (32,112,96,1) and each clinical data has dimension 22. For MRI data first we apply 2 consecutive 3D convolutions with kernel size (3,3,3) and the number of filters are 8 for first and 16 for second and sliding is one and keeping size of original remains same as convoluted matrix. Then Dimension of the convoluted matrix will be (32,112,96,16). Then we apply 3D max pooling with kernel size (2,2,2) on it, so the dimension will be (16,56,48,16). In this way i.e first apply convolution then max pooling finally we get a convoluted matrix whose dimension is (2,7,6,128) (see figure below).
- After getting a convoluted matrix we first do batch normalization then apply Global average pooling. Finally we get a feature matrix whose dimension is 128.
- We apply an autoencoder on clinical data and the loss function we have used is MME and the optimizer is ADAM. So the dimension of clinical features is reduced to 16, a latent space representation of clinical data.
- To get a joint representation we apply Multimodal deep framework technique on convoluted features of images and latent features of clinical data.
- Next we apply MLP on joint representation. We apply consecutively three hidden layers and the size of hidden nodes are 256,64 and 16 respectively. Then we apply softmax on the last layer where the number of hidden nodes is three. The loss function we have used is cross entropy (CE) and the optimizer is ADAM with more than 30 epochs for training.
- Finally we get three stages of classification on given MRI and clinical data.
- Before we train our model we initialize weights by Glorot weight initialize method. For avoiding overfitting we used the dropout method. For details discussion of above methods are given below

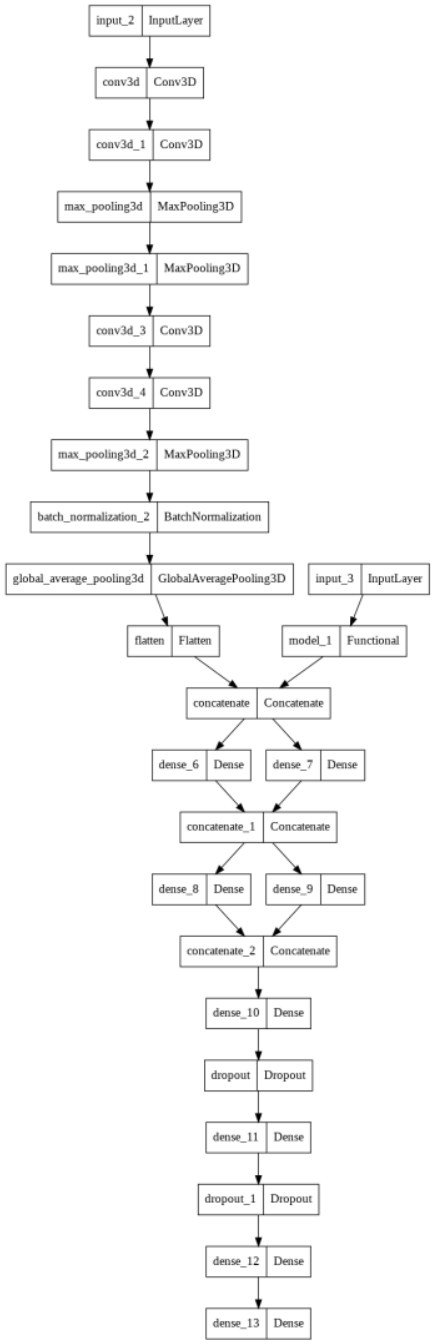

Figure 5: Direction of our proposed Architecture

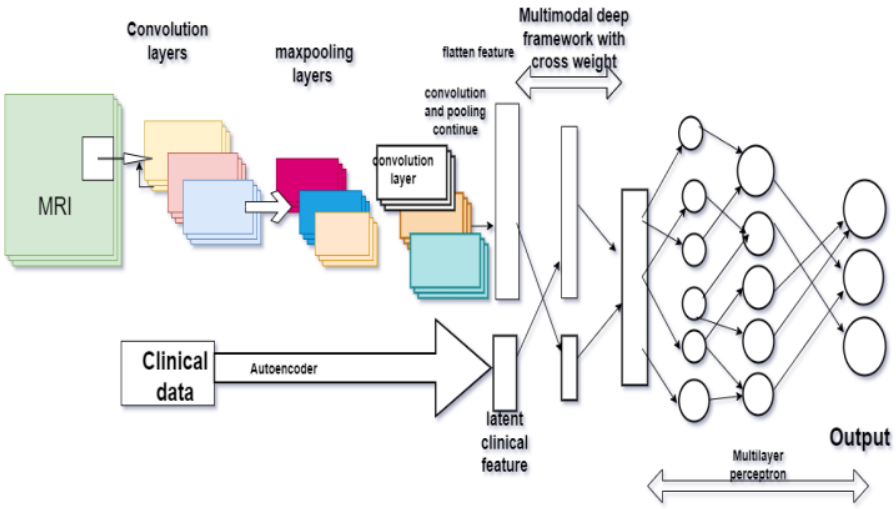

Figure 6: Our proposed architecture

3.1 Convolution

We use 3D convolution for extracting feature from MRI.
Convolution of a function $f(t)$ with a function (say weight function or kernel) $g(t)$ is defined by

$$f \star g = \sum_{u} f(u)g(u-t) \tag{1}$$

Similarly convolution on an image with some kernel is finding convoluted values in a small window of image selected by the kernel. Kernel sliding over the whole image and finding a convoluted matrix. Here we use zero padding on the boundary region to produce same size convoluted matrix of a image.

3.2 Global Average Pooling (GAP)

We use GAP after convolution and pooling layer on MRI data to get a column feature for each image.

Pooling is basically a down sampling technique.It is used for tensors . It takes the same values from each band of a tensor then averaging them and gives an array. The size of each array is the total number of bands. For an example given a tensor $T_{(x,y,z)}$ with dimension (12,16,20) after GAP its dimension will be 20. First it fixed z value then varying all x and y value finally averaging all values (see fig below).

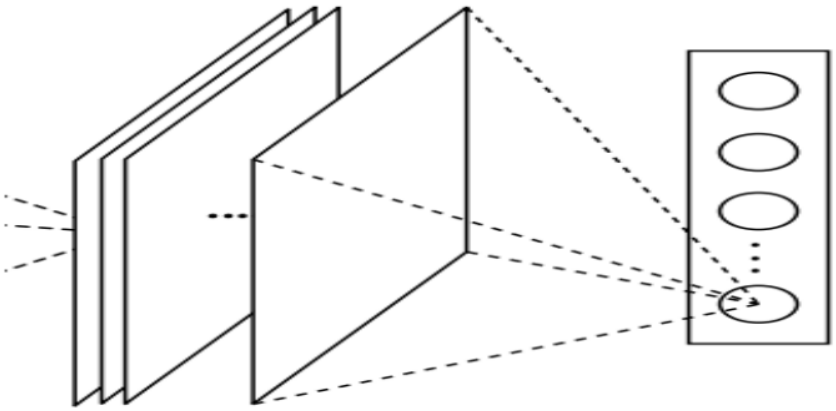

Figure 7: Our proposed architecture

3.3 Loss function

Loss function plays an important role for training any model. For better learning of weights of our model we need to give proper loss function. Loss function is depending on model to model. In our model we use two kinds of loss function and they are Mean Square Error (MSE) and cross entropy(BCE). The detailed discussion about this loss is given below.

3.3.1 Mean Square Error (MSE)

We use MSE loss for generating autoencoder on clinical data. MSE is defined by

$$L(y, y^{'}) = \frac{1}{N}\sum_{i=1}^{N} ||y - y^{'}||^2$$

(2)

where y is actual value of training set and $y^{'}$ is predicted value of any model.

3.3.2 Cross Entropy (CE)

We use CE loss function in our proposed architecture.
Cross Entropy is defined by

$$L_{CE} = -\sum_{i=1}^{C} T_i \log p(y_i) \tag{3}$$

where T_i is actual label and $p(y_i)$ predicted probability of ith class. In our model number of class is three i.e $C = 3$.
Average cross entropy for training dataset is defined by

$$L_{ACE} = -\frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{C} T_{ij} \log p(y_{ij}) \tag{4}$$

where T_{ij} is actual label and $p(y_{ij})$ predicted probability of ith class and jth data point respectively. Here N is total number of training points

3.4 Weight Initialization

Weight initialization is one of the important techniques for designing deep neural network models. Previously, weight initialization was done by using small random numbers. Last few years, some heuristics techniques have been developed . Some weight initialization methods are He initialization, Glorot initialization. In our proposed method we use Glorot weight initialization method which is discuss below.

Glorot Weight Initialization The Glorot initialization method [5] takes weights randomly with in uniform probability distribution on $[\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}]$ where n is the number of inputs to the node. The normalized Glorot initialization method takes weights (W) as a random number and defined as with a

$$W_{weight} = U[-\frac{\sqrt{(6)}}{\sqrt{(n+m)}}, \frac{\sqrt{(6)}}{\sqrt{(n+m)}}] \tag{5}$$

Where U is a uniform probability distribution and n,m are the number of input nodes and output nodes respectively (i.e. number of nodes in the previous layer to nodes in the current layer).

3.5 Autoencoder

We use an autoencoder for clinical data. Autoencoder is one type of dimension reduction technique from feature space to a latent space. It has two part encoder and decoder (see figure 8). Encoder encodes a given feature matrix, let say X to an encoded matrix Y by some function f (sigmoid, relu,tanh etc.) and weight W(see equation 6). Decoder decodes from encode feature matrix Y to decode matrix Z(equation 7).

$$Y = f(W^T X) \tag{6}$$

$$Z = g(W_1^T Y) = g(W_1^T f(W^T X)) \tag{7}$$

The loss will be calculated between X and Z by any loss function. Finally we get an encoded feature in Y.

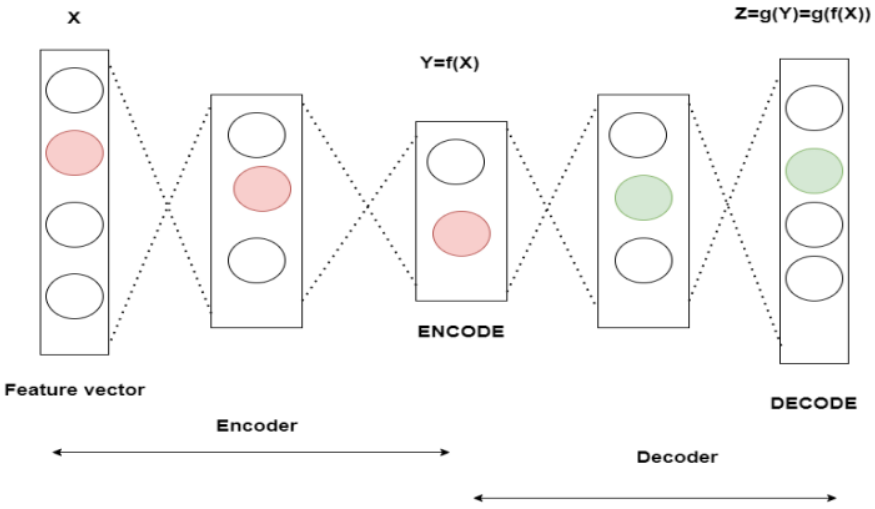

Figure 8: Autoencoder

3.6 Multimodal deep framework with cross weight

Multimodal deep framework[9] is a method to find a joint representation of two different modality features. Let X,Y be two feature vectors from different modality. Then the next step will be finding hidden layers individually for both by the cross weight(see fig 9). Let X1,Y1 be the first hidden layer of X,Y respectively. In terms of probability we

have to find the probability density function(PDF) of random variables of X,Y respectively. let f_{θ_1}, f_{θ_2} be probability density functions of X and g_{α_1} g_{α_2} probability density functions of Y. Then representation Random variables X1,Y1 are given below.

$$X1 = f_{\theta_1}(X) + g_{\alpha_1}(Y) \tag{8}$$

$$Y1 = f_{\theta_2}(X) + g_{\alpha_2}(Y) \tag{9}$$

In this way we find the next hidden layer. Final stage we will find joint PDF from previous layer of both.
Practically we will find these PDFs by deep belief networks. A connected layer from X and Y to X1 and another connected layer from X and Y to Y1 and continue in this way. The last stage has only one fully connected layer and weights will be trained from some loss function.
We use this method for joint representation of clinical data and MRI data.

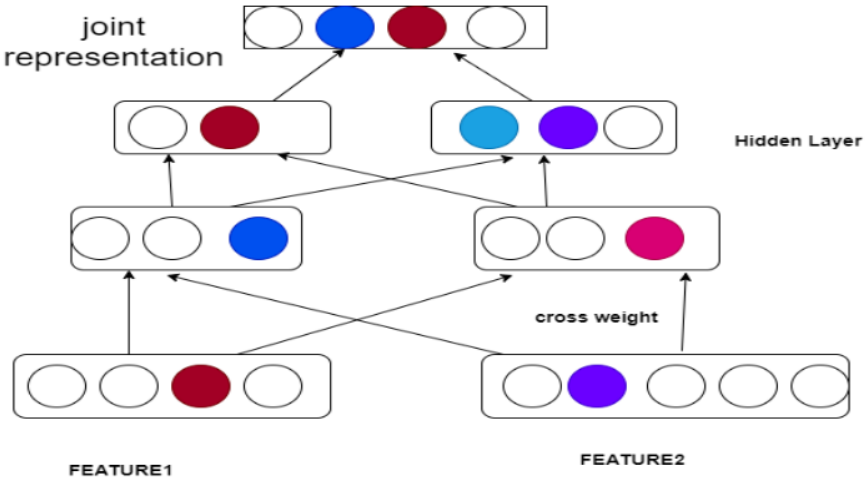

Figure 9: Multimodal data fusion method by deep framework with cross weight

3.7 Dropout

Dropout [13] is one kind of regularization method to resist overfitting of a model. We use dropout for training a model with a large number of neural networks (i.e network contains many hidden layers and nodes). If we put dropout in a layer then it randomly disconnects some nodes in that layer. So network architecture changes in that layer. Dropout uses hidden layers and input layers. It can not apply on the output layer.

Because of randomness for dropping node It prevents overfitting in our model. For example if the model is very much dependent on a particular feature or node and we apply a 50% dropout then chances of removing that node is in every two batches. Therefore other nodes are trained after every two batches.

3.8 Batch Normalization

Batch Normalization is used for boosting the training process , reducing generalization error and reducing the number of epochs in deep neural networks. We use batch normalization [6] after convolution layer and pooling layer in our model. During model training we take one batch at a time. Batch normalization is a popular technique. It standardizes inputs to a layer in a batch. we apply it either directly to the inputs or the output of the activation from the previous layer. It prevents the spread of distribution of the input layer during the weights update and weight will not change drastically. This has the effect of stabilizing and accelerating the training process of a network. Therefore our model performs very well after applying batch normalization.

4 CHAPTER 5: Result

In this part we apply different kinds of machine learning techniques on given data and compare them. The values of precision and recall corresponding to the AD stage.

In the First part we apply machine learning methods on clinical data. Here we split the dataset into 80% for the training, 20% for testing and validation accuracy is calculated by 5 fold cross validation. Let us discuss different parameters and hyperparameters of different supervised classification methods.

- K- nearest neighbor (KNN) : Taking value of nearest neighbor is 10 i.e $k = 10$
- Random Forest (RF) : We use 200 decision trees with loss function is gini index.
- MLP : We use two hidden layers. The numbers of Nodes in two hidden layers are 10 and 15 respectively and the optimizer we have used is ADAM.
- Logistic Regression We have used cross entropy as a loss function and the regularizer is L2.

Performance of different methods on clinical data are given below.

Table 2: Clinical Data

Performance	Logistic regression	KNN	RF	MLP
Accuracy	69.01%	61.36%	66.36%	70.03%
Precision	61.9%	75.6%	57.1%	66.66%
recall	61.9%	57.14%	67.9%	72.22%

Here we have done some experiment results on convoluted features of processed image data. We apply different machine learning algorithms the same as above and also hyperparameters are all the same as clinical data.
Performance of different methods on MRI data are given below.

Table 3: Convolution feature of MRI

Performance	Logistic regression	KNN	RF	MLP
Accuracy	50.53%	48.36%	56.03%	58.85%
Precision	45.84%	41.6%	61.1%	74.06%
recall	51.49%	47.14%	54.29%	56.42%

Here we first normal merge feature of clinical data then use different classifier and result is given below. For Logistic Regression and RF parameter and hyperparameter same as above and for MLP we add more hidden layers.

Table 4: Clinical feature and Convolution feature of MRI

Performance	Logistic regression	RF	MLP
Accuracy	48.73%	67.29%	54.27%
Precision	35.83%	54.16%	58.33%
recall	37.95%	50%	54.16%

- From the above table we see that normal concatenation gives deteriorate performance in compare to single modality (like clinical data).

Here we extract statistical features like F1,GLCM from MRI and find performance. Features taken from GLCM are dissimilarity, contrast, homogeneity, energy etc with directions are 0,45,90,135 degrees.Then we normally aid all feature (Clinical, Convoluted, Statistical feature).We are keeping all parameter and hyperparameter same as above for RF and Logistic Regression and for MLP we increase hidden layer(table 5).

Table 5: GLCM and F1 features performance

Method	GLCM and F1 statistic	concatenating feature of GLCM,Convolution,Clinical
Logistic Reg.	53.09%	67
RF	47.40%	58.60%
MLP	49.38%	65.66%

- From the above table we see that merging features (see table 5) reduces performance in comparison to single modality data (clinical data).

In final part we split the our data (Clinical and GM images) 85% for train and rest for the test and training data split into 85% for train and rest for validation.
First we individually find a convolutional autoencoder for MRI and autoencoder for clinical data with loss function MSE on training data. Then we merge these two encoded features by Multimodal deep framework (proposed data integration method for joint representation) with a loss function CE. Next we apply MLP on joint representation with a loss function is CE. In this architecture model performance is given below.

Table 6: Performance of Individually train different part of our model

Individually train different part of our model	Clinical data and MRI
Accuracy	62%

- In this case performance above model is very poor.

Finally the performance of our proposed model is given below.

Table 7: Performance of our proposed model

proposed Method	Clinical data and MRI
Accuracy	$75^{+}_{-}5\%$
Precision	76%
Recall	71%

If we change loss function of proposed model then performance is slightly decreased. We have done some experiment by using loss KL divergence instate of CE then performance is given below.

Table 8: Performance of proposed method with loss function is KL Divergence

proposed method with loss KL Divergence	Clinical data and MRI
Accuracy	69.01%

superiority performance : We see that loss function CE gives best performance in comparison to KL divergence. We also see that our model gives better performance in comparison to individual training each part of our model with a loss function for each part. Proposed architecture gives good performance in comparison to other methods which are discussed above.
Finally we give a graphical picture how training accuracy is increasing and how loss is decreasing in our proposed model.

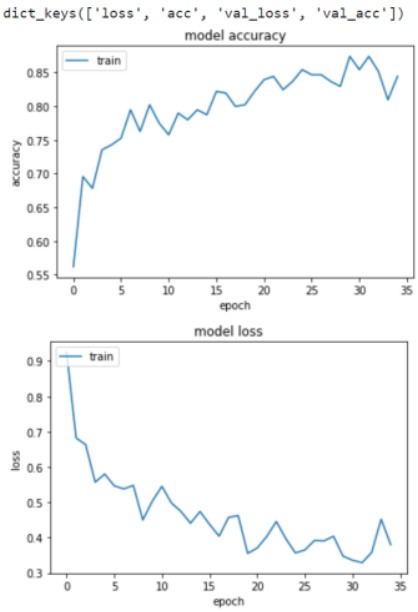

Figure 10: Graphical representation of training process

5 CHAPTER 6: Conclusion

Prediction of different stages of AD is very challenging. Only convolution on MRI can't learn very well on MCI stage. For multiclass problems we need very complicated models .If we have only two classes (CN,AD) then accuracy will be above 90% for any model. From the above results we see GLCM feature are not helpful for classification. performance of MLP and RF are better than other shallow model.
In this work our proposed method outperform any other model.
• Convolution is helpful for different AD stages analysis.
• Multimodal deep framework outperform Single modality.
• Performance of Normal feature addition of two modality is not well. Although our model suffers for small dataset if we get large dataset then Proposed method perform very well and gives very high accuracy.

References

[1] Mohammad R Arbabshirani, Sergey Plis, Jing Sui, and Vince D Calhoun. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, 145:137–165, 2017.

[2] Rémi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehéricy, Marie-Odile Habert, Marie Chupin, Habib Benali, Olivier Colliot, Alzheimer's Disease Neuroimaging Initiative, et al. Automatic classification of patients with alzheimer's disease from structural mri: a comparison of ten methods using the adni database. *neuroimage*, 56(2):766–781, 2011.

[3] Sandrine Cure, Keith Abrams, Mark Belger, Michael Happich, et al. Systematic literature review and meta-analysis of diagnostic test accuracy in alzheimer's disease and other dementia using autopsy as standard of truth. *Journal of Alzheimer's Disease*, 42(1):169–182, 2014.

[4] Farshad Falahati, Eric Westman, and Andrew Simmons. Multivariate data analysis and machine learning in alzheimer's disease with a focus on structural magnetic resonance imaging. *Journal of Alzheimer's Disease*, 41(3):685–708, 2014.

[5] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[6] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[7] Alexander Khvostikov, Karim Aderghal, Jenny Benois-Pineau, Andrey Krylov, and Gwenaelle Catheline. 3d cnn-based classification using smri and md-dti images for alzheimer disease studies. *arXiv preprint arXiv:1801.05968*, 2018.

[8] Dougho Park, Byung Hee Kim, Sang-Eok Lee, Dong Young Kim, Mansu Kim, Heum Dai Kwon, Mun-Chul Kim, Ae Ryoung Kim, Hyoung Seop Kim, and Jang Woo Lee. Machine learning-based approach for disease severity classification of carpal tunnel syndrome. *Scientific Reports*, 11(1):1–10, 2021.

[9] Sarah Rastegar, Mahdieh Soleymani, Hamid R Rabiee, and Seyed Mohsen Shojaee. Mdl-cw: A multimodal deep learning framework with cross weights. In

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2601–2609, 2016.

[10] Saima Rathore, Mohamad Habes, Muhammad Aksam Iftikhar, Amanda Shacklett, and Christos Davatzikos. A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer's disease and its prodromal stages. *NeuroImage*, 155:530–548, 2017.

[11] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps*, pages 323–350, 2018.

[12] Alessia Sarica, Antonio Cerasa, and Aldo Quattrone. Random forest algorithm for the classification of neuroimaging data in alzheimer's disease: a systematic review. *Frontiers in aging neuroscience*, 9:329, 2017.

[13] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[14] Nitish Srivastava and Russ R Salakhutdinov. Multimodal learning with deep boltzmann machines. *Advances in neural information processing systems*, 25, 2012.

[15] Srikanth Tammina. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, 9(10):143–150, 2019.

[16] Muhammad Tanveer, Bharat Richhariya, Riyaj Uddin Khan, Ashraf Haroon Rashid, Pritee Khanna, Mukesh Prasad, and CT Lin. Machine learning techniques for the diagnosis of alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(1s):1–35, 2020.

[17] Janani Venugopalan, Li Tong, Hamid Reza Hassanzadeh, and May D Wang. Multimodal deep learning models for early detection of alzheimer's disease stage. *Scientific reports*, 11(1):1–13, 2021.

[18] Michael W Weiner, Dallas P Veitch, Paul S Aisen, Laurel A Beckett, Nigel J Cairns, Robert C Green, Danielle Harvey, Clifford R Jack, William Jagust, Enchi Liu, et al. The alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's & Dementia*, 9(5):e111–e194, 2013.

AMultimodal Deep Learning Framework for Prediction of Alzheimer's Disease

ORIGINALITY REPORT

5%
SIMILARITY INDEX

PRIMARY SOURCES

1	library.isical.ac.in:8080 Internet	151 words — 3%
2	link.springer.com Internet	66 words — 1%
3	Palak Goyal, Rinkle Rani, Karamjeet Singh. "State-of-the-Art Machine Learning Techniques for Diagnosis of Alzheimer's Disease from MR-Images: A Systematic Review", Archives of Computational Methods in Engineering, 2021 Crossref	21 words — < 1%
4	www.arxiv-vanity.com Internet	15 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 14 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 14 WORDS