A Novel Deep Learning Model for Prediction of Alzheimer’s Disease Based on Neuroimaging

[[1]](#footnote-1)

***Abstract*****: Computerized health care has grown rapidly due to advances in medical imaging and machine learning technologies. In particular, recent advances in deep learning are ushering in a new era of clinical decision-making based primarily on the multimedia system. Alzheimer's disease (AD) is identified as impaired psychological traits and severe amnesia. These changes occur in brain structures due to shrinkage of gray and white matter of brain and also many more reasons. It can be measured using magnetic resonance imaging (MRI) scans; these scans provide an opportunity for prior detection of AD using classification tools such as Convolutional neural network (CNN). In any case, as of now, most AD-related tests have been restricted by test measure. Finding a cost-effective way to train image classification on limited knowledge is important. In the proposed work, we studied distinctive transfer-learning techniques based on CNN for AD prediction using brain structure MRI scanning and selected and improved the best technique for better accuracy.**

**Index terms: Alzheimer’s disease, CNN, MRI, Amnesia.**

# Introduction

Like all the world's great populations, India's huge population (the second largest in the world) is facing a crisis between the elderly and the infirm. The crisis is that Alzheimer's terrible degenerative disease that will abruptly and persistently affect anyone at any time until it takes on the victim its final, terminal toll. More than 4 million people in India have some kind of dementia. A minimum of 6 million people worldwide are living with dementia, making the disease a global health crisis that needs to be addressed.

Alzheimer’s disease is the most typical reason for dementedness. The symptoms of AD is chronic brain disorder of dementia that includes amnesia and difficulties with thinking, problem-solving or language that have dire effect on a patient’s lifestyle. AD is a chronic neurodegenerative disease that typically starts slowly and gets worse over time. Moreover, the reason behind AD is poorly understood. No treatments stop or reverse its progression, although some might briefly improve symptoms.

Till date, AD is mostly detected at a late stage at which treatment will solely slow the progression of cognitive decline.

In order to improve preventive and disease-modifying therapies, early detection of AD is very important. At the onset of Alzheimer's disease, individuals may suffer from Mild Cognitive Impairment (MCI), an intermediate stage between the expected decline in traditional aging psychological features and the additional severe decline in

dementia. It implies that the brain has a mild cognitive and memory impairment, but it has no effect on the daily functioning of the individual and can hardly be detected in clinical applications.

Previous research has found that the risk of AD plague with MCI is greater than that of traditional people [1],[2]. The prevalence rate of individuals with MCI is based on an annual rate of 100 percent [3] and the traditional old people are eighteen-two per year [7]. Many machine learning strategies applied to structural imaging [1] have been used by computer-aided classification of AD and MCI patients. Support Vector Machine is the most popular among these strategies. SVM extracts high-dimensional, informative imaging options to create prognosticative classification models that facilitate clinical designation automation. Definition of feature extraction, however, generally believe manual extraction from semi - automatic brain structure outline, which is toiling and at risk of inter- and intra - rater variability, or complicated pre - processing of images, which is long and computationally difficult to please.

An alternative family of machine learning strategies, referred to as deep learning algorithms, achieve optimal results in many areas such as tasks of speech recognition, computer vision and understanding of natural language (Lecun et al., 2015) and, more recently, medical analysis [5]. Deep learning algorithms differ from conventional machine learning methods because they require little or no pre-processing of images and can automatically infer optimal representation of data from raw images without requiring prior selection of features, resulting in a process that is more objective and less bias-prone [6]. Deep learning algorithms are therefore better suited to detecting subtle and diffuse anatomical abnormalities. Recently, to identify AD patients from normal controls, deep learning has been successfully applied to the Alzheimer's disease Neuroimaging Initiative (ADNI) dataset. Until now, only one report has linked profound learning calculations without the earlier determination of elements (considering dim issue [GM] volumes as contribution) to the expectation of AD advancement within one and a half years in people with MCI using ADNI auxiliary MRI controls. Several machine learning approaches [8], [9], [10], [11] have recently been used to obtain pathological biomarkers for the diagnosis of multi - modal

neuroimaging knowledge supported by AD / MCI, along with resonance imaging (MRI), positron emission imaging (PET), etc.

In the recent decade, the convolutional neural network has been widely used for image classification tasks with glorious performance. While, a well-performed CNN image classifier, e.g. AlexNet and ResNet, is typically developed to support an enormous quantity of training data, that is impractical for medical image classification, because of restricted resource, particularly brain tomography.

Fine-tuning a neural network using transfer learning [6] is much more convenient than to train a network from scratch. Trained CNNs are constructed by precisely training the CNN on large-scale datasets which can be later used for Image Processing applications. Then these CNN’s are used in the subsets of their respective image processing domains using Transfer Learning, where only the last layer needs to be precisely adjusted last layer to the need of the new subset CNN. Transfer learning are networks trained on natural pictures used with medical images, and are verified to be robust even for cross-domain applications [8]. Therefore, CNN connected strategies are suitable for learning from a small set of low-scale training can be tremendously helpful in developing a predictive AD classifier using tomography image. Transfer learning is one of the possible solutions. The idea of transfer learning is to pre-train a ConvNet on a really massive dataset (e.g. ImageNet), then use the ConvNet for the task of interest either as initialization or as a set feature extractor.

# literature survey

Jin Liu, Min Li, Wei Lan, Fang-Xiang Wu, Yi Pan, and Jianxin [1] Wang proposed a model that predicts MCI which can convert into AD. Firstly, they have selected some regions based on AAL (Automated Anatomical Labeling) which is a software and a digital human brain atlas with a labeled volume that is making a map of the brain and giving different regions some name Jin Liu, Min Li, Wei Lan, Fang-Xiang Wu, Yi Pan, and Jianxin [1] Wang proposed a MCI model that could be converted to AD. First, they selected some regions based on AAL (Automated Anatomical Labeling), which is a software and a digital human brain atlas with a volume labeled that makes a map of the brain and gives a name to different regions. Labels indicate macroscopic brain structures from the MRI images and then build entire brain hierarchical network where a hierarchy of their region of interest is present and on the basis of finding the strength of the connection between regions. In terms of Pearson's correlation coefficient, the connectivity between each pair of regions is calculated and used as a classification feature. The selected the features with higher F-scores are used to reduce the dimensionality of the features. Finally, the classification is performed using multiple kernel boosting (MKBoost) algorithm.

Ronghui Ju, Chenhui Hu, Pan-Zhou, Quanzheng Li [3], used resting-state fMRI data for early detection of Alzheimer’s disease. The brain is divided into 90 regions and the R-fMRI data is transformed into a 90 × 130 matrix which retains the primary information. Pearson’s correlation coefficient is used to measure the strength of the links. Based on the correlation coefficient, the time series data is transformed into a 90 × 90 correlation coefficient matrix and a complete functional brain network is constructed. The correlation coefficient data is the basis for detecting MCI. In addition, the clinical examination data (including age, gender, and genetic information) help to analyze the relationship between MCI and other physiological factors. Then, a deep autoencoder network model is built to categorize these correlation coefficient data. Deep learning model based on stacked autoencoders that have been developed to extract hierarchical features in high-dimensional data.

Tong Tong, Qinquan Gao, Ricardo Guerrero, Christian Ledig [4], Liang Chen have developed an effective biomarker for an accurate prediction of MCI-to-AD conversion from magnetic resonance (MR) images. They have proposed a novel grading biomarker for the prediction of MCI-to-AD conversion. Firstly, they have comprehensively studied the effects of several important factors on the performance in the prediction task including registration accuracy, age correction, feature selection and the selection of training data. Based on these factors, a grading biomarker is then calculated for each MCI subject using sparse representation techniques. Finally, the grading biomarker is combined with age and psychological feature measures to produce correct prediction of MCI-to-AD conversion.

F. Barkhof, S. Haller, and S. A. R. B. Rombouts [5] used system Resting-state (RS) purposeful MR imaging that overcomes the restrictions of task-based MR imaging by searching multiple vegetative cell networks at the same time throughout a 5–10-minute acquisition and divulges brain physiology. Data analysis techniques are still evolving from a simple region of interest–based correlation analyses to data-driven methods, graph theory, and pattern recognition. Neurologic and psychiatric diseases are often characterized by complex alterations in the pattern of multiple functional networks, not only by single networks such as the default mode network.

Gang Guo, Min Xiao, Min Du, Xiaobo Qu [9] proposed an approach structured on (CNN), and is made to accurately anticipate MCI-to-AD transformation using magnetic resonance imaging (MRI) information. Initially, MRI images are processed with age-correction. Next, regional areas, which have been constructed in 2.5 proportions, tend to be produced from these images. Then, these areas utilized in order to train the CNN to find the MCI subjects. Subsequently the, brain image features were excavated with free Surfer in order to boost CNN. Lastly, both the types of features were supplied with into an intensive ML classifier to predict AD.

Marcia Hon, Naimul Mefraz Khan [10], attempted to solve some basic constraints like depending upon extensive variety of training images and also the demand for properly boosting the structure of CNN and tried to discover such issues among transfer learning, when advanced architectures such as VGG are initialized using pre-trained weights from massive standard datasets that consist of normal images, as well as the fully-connected layer is trained again with just a tiny amount of MRI images. They utilize graphic degeneration to select the highest beneficial slices for training degeneration to select the highest beneficial slices for training. But its major drawback was lack of control over the Convolutional layers.

# Method

## Architecture

## The basic architecture for this project consists of three models. Firstly the pre-trained AlexNet model which is trained on MRI scans of different disease related to brain like dementia, tumor, Encephalitis, Parkinson, Alzheimer’s etc. This is done due to the lack to training dataset available for us to train our machine. In the second phase the features of existing Pre-trained model is transferred to a new AlexNet model. In the third phase, images are taken from ADNI dataset to train and fine-tune our model which is designed specifically for Alzheimer’s disease prediction. The process is described in detail in the later part of the paper.

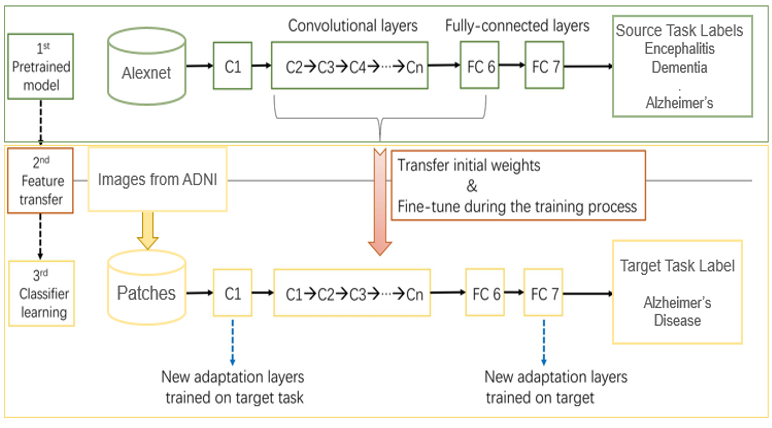


Figure 1: System Architecture

## Data Acquisition

In this research, the MRI scans of Alzheimer’s disease Neuroimaging Initiative (ADNI) [1] is utilized to prepare the convolutional neural network (CNN) classifier. ADNI is actually an in development, adherent study that was started in 2004. It aims on comprehending the diagnostic and prognosticative terms of Alzheimer's disease-specific biomarkers. The information enclosed an overall 715 structured MRI scans out of both ADNI1 as well as ADNI2 phases, having 320 AD cases and 395 normal controls. The basic idea was to arbitrarily evolve the entire sample into training dataset, validation and testing dataset. Then the data is acquired for dataset with the previously mentioned description in the DICOM MRI format. For further investigation of the information it required to convert the data into NIfTI format so that we can further work on it and provide it to the accessible neuroimaging toolbox, where the further preprocessing is applied to the data.

## Data Preprocessing

The raw image data needs preprocessing of images for that Statistical Parametric Mapping (SPM) is used. SPM is basically a statistical technique which utilized for evaluating cognitive activities documented throughout neuroimaging tests. The statistical parametric mapping software system is additionally used for spatial standardization, smoothing, and statistical analyses of the parametric images. The original magnetic resonance images were initially stripped of skull and segmented using the segmentation algorithm based on the probability mapping and then standardized using affine registration to the International Brain Mapping Template. The setup comprises of bias, noise and overall intensity calibration. The traditional way of preprocessing will create image records having 122X145X122 standardized size. The removal of the skull and normalization offers equivalence within the images simply by adjusting the brain's initial image into a normal image space, positioning the same brain substructures with the image's coordinates alongside similar image of different participants.

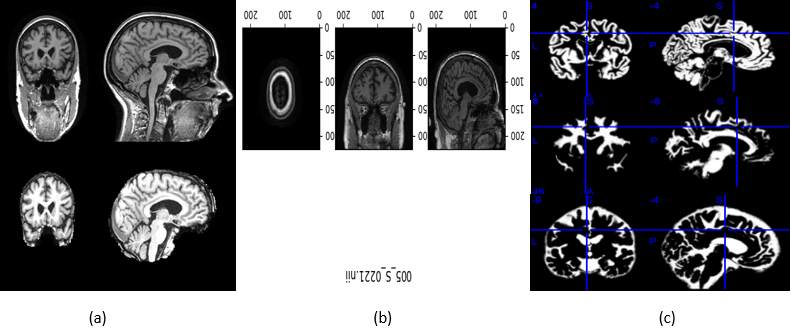


Figure 2: Image pre-processing: (a) Skull stripping, (b) Template matching, (c) White matter, Grey matter, csf extraction and selection.

## Our Experiment

A CNN (Convnet) is a type of neural network consists of several layer which are: input layer, several hidden layer and output layer. The hidden layer is the most important layer it is used for feature extraction. It comprises of Convolution layer

(detects pattern), Rectified Linear Unit (ReLU) (used to get rectified feature map), pooling layer (uses filters to detect edges etc.) and fully connected layer. And another main advantage of ConvNet is that is assumes that the input data is image only that helps to embed certain properties into it.

## AlexNet

In this research, AlexNet Architecture has been used as a CNN classifier. AlexNet is a popular and precise CNN architecture which made its name in machine learning and AI notably in Image processing. The basis advantage of this architecture is that it has pooling layer after every Convolutional layer and we can modify the Convolution layer as it doesn’t have fixed size of convolution layer which gives us more control over the CNN. It consisted of 11X11, 5X5, 3X3, convolutions, max pooling, dropout, data increase, activation of ReLU, dynamic SGD. After each convolutional and fully connected layer, it is attached to ReLU activations.

We looked at two different architectures, AlexNet and GoogleNet, because they were successful on similar issues [10], [11]. AlexNet was at par with GoogleNet. In both of the trained architectures, 25000 random images of test dataset were used to compare the accuracy of the two architectures. Then the confusion matrices and the accuracy for each test dataset are calculated for both the architectures.

AlexNet's input is a 224x224x3 RGB image that passes through the first convolution layer with 96 feature maps or filters with a size of 11X11 and a 4 stride. The size of the image changes to 55x55x96. Then the AlexNet applies a maximum layer of pooling or sub-sampling layer with a filter size 3 underground and a step of two. Dimensions of the resulting image will be reduced to 27x27x96.

The primary variation in our AlexNet versus the standard AlexNet comes in second layer. In the 2nd convolution layer it receives data coming from the 1st convolution layer then screens it with 256 feature maps of dimension 5X5X64. Subsequently the next convolution layer is connected to the data coming from 2nd convolution layer by 384 feature maps

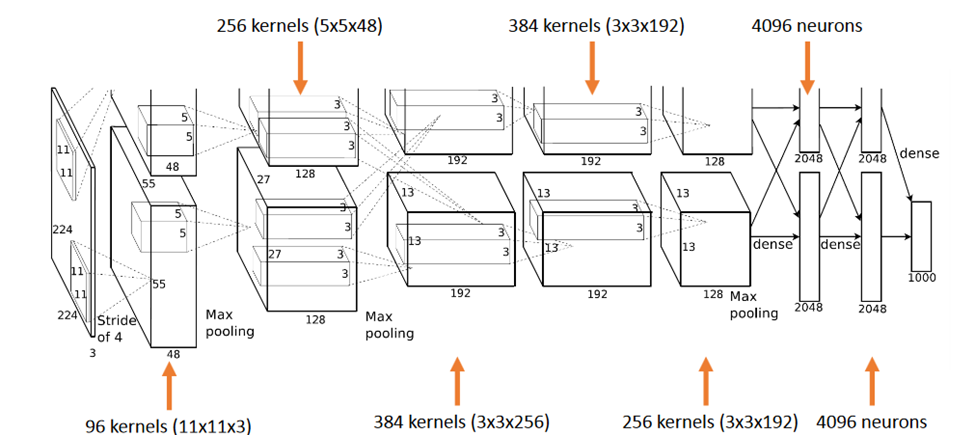


Figure 3: AlexNet Architecture

of dimension 3X33X192. The next layer of convolution has 384 feature maps of dimension 3X3X84, and the 5th layer of convolution has 256 feature maps of dimension 3X3X56. These are the ideal dimension for feature extraction purpose in these types of images and in order to train them properly. In sixth layer the output is flattened through a fully connected layer in order to accommodate all the data coming from AlexNet Architecture, we have done some minute adjustments, a fully connected layer is included with 985 neurons in the base model with a softmax layer to deal with negative loss.

## Activation function

An activation function is the function in an artificial neuron that delivers an output based on inputs. In AlexNet we are using ReLU as activation function Instead of Tanh and sigmoid, ReLU activation function is used to add non - linearity.

Hyperbolic tangent function:

(1)

Sigmoid function:

(2)

These functions are slow to train.

The advantage of using the ReLU over other activation functions is that it doesn't activate all the neurons simultaneously. Meaning If you look into the ReLU operate if the input is negative it converts it to zero and therefore the neuron doesn't get activated. Rectified Linear Unit (ReLU) is faster to train where:

(3)

It speeds up the speed at the same accuracy by 6 times.

## Fine tuning

In this half, the optimizer we tend to use is the Adam optimizer, one of these algorithms that work well across a wide variety of deep learning architectures is the Adaptive

Moment Estimation or Adam optimization algorithm. Adaptive Moment Estimate (Adam) [14] calculates for each parameter adaptive learning rates. It is also referred as solver. The declining averages of past and past square gradients mt and vt are calculated as follows:

(4)

(5)

mt and vt are estimates of the first moment (the mean) and the second moment (the uncentered variance) of the gradients respectively, hence the method name. By calculating bias-

correction, first and second moment estimates, they counteract these biases:

(6)

(7)

We use these to update the parameters that give the update rule for Adam:

(8)

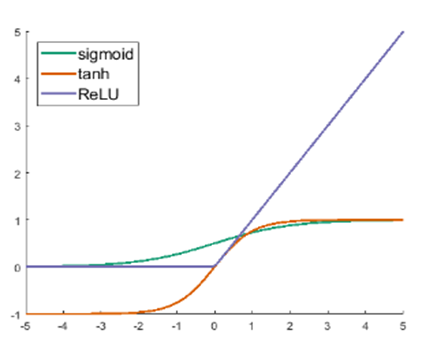
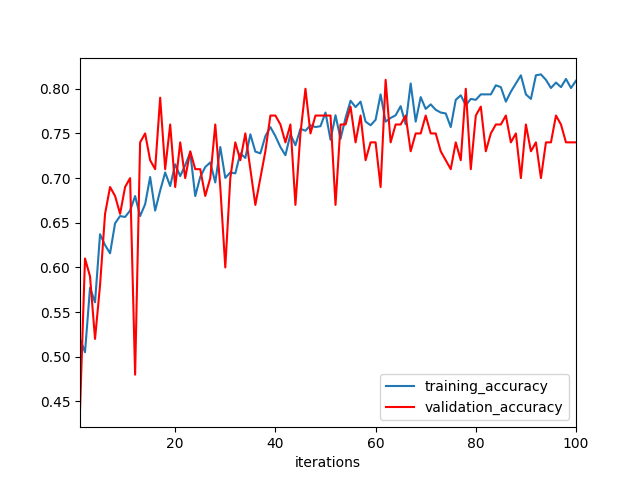


Figure 4: Activation Function

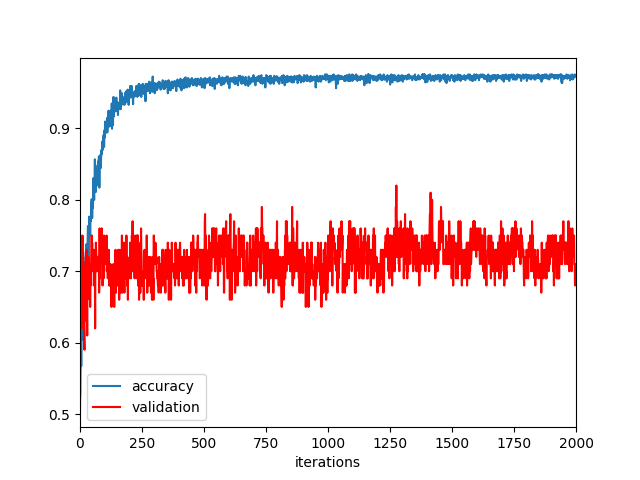
## Fine-tune the linear fully-connected layers

As we know the network's end layers tend to be much more precise to the class features in the initial dataset and the pretrained AlexNet dataset vary from the initial dataset. So, in order to tackle this situation we have to replace the last classifying layer, and also calibrate all the classifier. For solving this problem first fully-connected layer is back propagated and its weights are kept in earlier convolution layer.

Calibrating the final layer of convolution and the classifiers formulated on the observation that the networks end layer contains more precise characteristics for the dataset. We cannot discover that only calibrating it for the fully connected layer will solve all the problems. Subsequently, we have carried out some trials to adjust the final output layer by the use of the last convolution layer. Mainly because, contorting with the initial parameters can be disastrous for the network and our research, the learning rate or we can say weight controlling parameter that we have used is 1e−4 the learning rate is increased for enhanced learning and fine-tuning, which is showing better performance than the standard learning rate.



(a)



(b)

Figure 5: Analysis of learning rate: (a) With learning rate 1e-3, (b) With learning rate 1e-4

In our experiment with learning rate 1e-4 max validation accuracy achieved is 80% and max training accuracy is 85% and when we increase the learning rate to 1e-4 we achieved achieved max validation accuracy as 82% with corresponding

training accuracy as 96.5% and max training accuracy is 97.5%.

Additionally, as mentioned earlier the improper initialization of the network parameter could impact the performance of the network. But to set uniformly the initial parameter there are various algorithms available and we have used Xavier Initialization algorithm based on some research. SVM is used for the final classification and AlexNet as an extractor.

## Transfer learning

Small data set limits the use of complex neural network as it would overfit the training data, while transferring learning shows good performance by using a pre-trained network (e.g. AlexNet) to extract features or fine-tuning parameters. After finishing the AlexNet's higher layer, we achieved 86 percent accuracy in the AD / Normal classification. We also find that using Xavier initialization could achieve a slightly better performance compared to random weights that many previous works have revealed [19] (By using Xavier initialization, we ensure that the weights are not too small but not too large to accurately propagate the signals. Initialization is important. For a small network with limited number of layers, a significant variance can be there while initializing the weights for distribution having zero mean and variance:

(9)

where nin and nout are respectively the number of inputs and outputs in the layer.

To perform classification using SVM, we performed the experiment using different layers of output features. Previous studies [11-13] suggested that generic characteristics are usually extracted from lower layers. Our results indicated that these generic features are not good for SVM classification (the accuracy is 55 %), while the features in the higher layers contain more specific characters that can be distinguished by a linear SVM (the accuracy is 85 %).

# Conclusion

The aim of this research is to classify Alzheimer's disease (AD) using transfer learning. We used strict pre-processing steps on raw MRI data from the ADNI dataset and used the AlexNet ie. CNN classifier on pre-processed data to is used to classify Alzheimer’s disease. During classification, features of low to high levels were learned. We increased the learning rate for enhancing the prediction of our model. For future classification of neural networks, pre - trained convolution layers capable of extracting generic image features such as pre - trained AlexNet convolution layers on CNN could provide good input features.

# References

1. Alzheimer’s Disease Neuroimaging Initiative, http://adni.loni.usc.edu/.
2. Ronghui Ju, Chenhui Hu, Pan Zhou, Quanzheng Li, “Early Diagnosis of Alzheimer’s Disease based on resting-state brain networks and deep learning”, IEEE transactions on computational biologyand bioinformatics, vol 64, no 2, September 2018.
3. M. C. Carrillo, “2014 Alzheimer’s disease facts and figures,” Alzheimer’s Dementia, vol. 15, no. 4, pp. e53–e82, 2018.
4. Li, Hongming, et al. “A Deep Learning Prognostic Model for Early Prediction of Alzheimer’S Disease Based on Hippocampal MRI Data.” Alzheimer's & Dementia, no. 7, 2018.
5. Tong Tong, Qinquan Gao, et al“A novel Grading Biomarker for the Prediction of Conversion From Mild Cognitive Impairment to Alzheimer's Disease”. IEEE Transactions on Biomedical Engineering Volume: 64, Issue: 1, Jan. 2017.
6. S. Liu, Y. He, “Regional homogeneity approach to fMRI data analysis,” NeuroImage, vol. 22, no. 1, pp. 394–400, 2017.
7. Helena Barros , et al, “Atlas based sparse logistic regression for alzheimer’s disease Classification” 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2017.
8. J. Lium Li , C. Wee, P. Yap, “A survey of MRIbased brain tumor segmentation methods,” Tsinghua Sci. Technol., vol. 19, no. 6, pp. 578–595, 2014.
9. N. L. C. Talbot*,* G. Chen, B. D. Ward, “On over-fitting in model selection and subsequent selection bias in performance evaluation,” Journal of Machine Learning Research, vol. 12, pp. 2059-2107, 2015.
10. F. Barkhof, S. Haller, and S. A. R. B. Rombouts, “Resting-state functional MR imaging: a new window to the brain,” Radiology, vol. 272, no. 1, pp. 29-49, 2014.
11. Y. Chao-Gan, Z *et al*, “DPARSF: a MATLAB toolbox analysis of resting-state MRI,” Frontiers in systems neuroscience, vol. 4, 2014.
12. Dragan Gamberger, Bernard Z Enko, Alexis Mitelpunkt, Netta Shachar Nada Lavrac, “Clusters of male and female Alzheimer’s disease patients in the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database.” Springer 2016.
13. Lavneet Singh, Girija Chetty, “A comparative study of MRI data using various machine learning and pattern recognition algorithms to detect brain abnormalities”, Australian data Mining Conference, 2017.
14. Dragan Gamberger, Bernard Z Enko, Alexis Mitelpunkt, Netta Shachar Nada Lavrac, “Clusters of male and female Alzheimer’s disease patients in the Alzheimer’s Disease Neuroimaging Initiative (ADNI)database. ”,Springer 2016.

1. [↑](#footnote-ref-1)