DEEP LEARNING NEURAL NETWORK APPROACHES FOR THE DETECTION OF ALZHEIMER'S DISEASE IN EARLY STAGE

*Nithya V. P, *Mohanasundaram N.*

Abstract

Alzheimer's Disease can be a primary cause for dementia in aged people. MCI - Mild Cognitive Impairment which is another sort of dementia. It is a prior phase of AD. It especially concerns aged populations. There is no proper medication and drugs to treat AD. Therefore, early detection of AD is necessary. There are couples of brains imaging strategies like MRI, PET and FMRI those can be used to realize various abnormalities in brain. Ongoing studies explain that deep models are more appropriate for prediction of AD in contrary to conventional AI techniques. This paper discussed about different Deep Learning (DL) techniques for prior identification of AD.

Keywords : Alzheimer's Disease(AD), Mild Cognitive Impairment(MCI) , MRI, PET, deep learning, Auto encoder (AE).

I. INTRODUCTION

AD is one of the reasons for dementia, but it is rare under the age of 45. The clinical feature is inability to remember new information [1]. AD is a baseline reason for dementia development in people and it causes a continuous decrease in thinking, behavioral and social abilities that interrupt a person's capability to feature independently. The early symptoms of the disorder may forget recent events or conversations. As the progress of Alzheimer's' disease affects a person with excessive memory impairment and loses the capability to perform normal tasks. There is huge contraction of cerebral cortex with AD, fatty blood vessels deposition and atrophied brain cells. Senile plaques and neuro-fibrillary are some common indications of AD. By determining the severity of disease, a physician performs prediction of disease symptoms those are considered in future for providing treatment guidelines.

Starting stage of Alzheimer 's' disease is characterized as MCI(Mild Cognitive Impairment) [2], where every patient with MCI is not developed with Alzheimer disease. In MCI, someone has slight cognitive variations are encountered among the individual subject and their relatives; however regular activities can be performed. Based on the statistics, roughly 15- 20% of aged people may have Mild Cognitive Impairment and 30% to 40% individuals show Mild Cognitive Impairment increase AD in five years. This changeover period ranges between 6 to 36 months however commonly 18 months. Patients are classified as MCI non-convertors (MCInc) or MCI converters (MCIc), which means patient may or may not converted to AD within 18 months. The most massive hazard elements are family histories and associated genes occurrence in an individual genome. Sporadic AD is a common type of AD. This is because of complex genes aggregation, environment and their life style. The main aspect for growing sporadic AD is getting old. Some cases may encounter it at the age of 60-65. There is no single test to find whether a person has AD or not. A diagnosis can be obtained from the existence of particular indications and identifying the reasons of dementia. It needs a detailed evaluation, based on medical history also requires, mental status examination, a physical exams, neurological exam, blood sample tests and brain image analysis.

Neuro-imaging is a crucial research direction for AD prediction. There are huge brain imaging process those are applied to recognize abnormalities over brain, consisting of

Department of CSE, Karpagam Academy of Higher Education
Coimbatore, Tamilnadu, India
*Corresponding Author

PET, short form of Positron Emission Tomography, MRI, short form of Magnetic Resonance Imaging, and CT scans, short form of computerized tomography. Some technique is related with each scan and it identifies specific structures and abnormalities in brain. Combining these methods increases classification accuracy. MRI is widely available and biomarker for AD prediction with superior performance. It uses radio frequency pulses and magnetic field for creating 3D specification of soft tissues, organs and bones. MRI is better compared to PET. This paper reviews the different deep learning (DL) methods utilized for AD prediction. DL techniques are more suitable compared to conventional machine learning (ML) techniques. ML techniques are composed of three preliminary phases and they are dimension reduction, feature extraction and classification. All these steps are combined while using DL methods. DL is a part of ML in AI applications that works alike human brain in processing data and recognizing pattern for finding solution in complicated decision making issues. DL techniques have applications in several areas, like image segmentation, object detection, tracking, and classification. Recent years, in medical imaging for disease detection, DL models, like Convolutional Neural Networks (CNNs) outperforms other techniques. With neuro-imaging data, DL model identifies hidden representations, association among various image parts, and recognizes disease based patterns.

II. PREPROCESSING MRI DATA

Preprocessing is required over the MRI data used for AD detection. There are lots of preprocessing steps. The preprocessing will be successful when it is an effective one. The preprocessing methods like registration, intensity normalization, skull stripping, tissue segmentation and motion correction can be used. Similarly, some DL strategies are anticipated with diverse pre-processing procedures.

In intensity normalization, intensity mapping of every voxel or pixel over reference scale is completed. Intensity normalization is compared with systems of same intensities. Some general techniques are non-parametric and non-uniform intensity normalization algorithm. Another technique like smoothing with Gaussian filter and Full-width at half maximum reduces noise level as preserving signal level. In registration, the images scans are spatially align to reference anatomical space. It is important because complex brain structures are large. Image registration attempts in neuro-imaging modalities standardization with a fixed-size template. It is probable to evaluate brain scans' voxel intensities of different individuals with their alignment, while fulfilling complete voxel in a single scanning with anatomic positions as inside some other patient's brain.

The volume of tissue in each portion of MRI is measured by the tissue segmentation method. Since, the grey matter (GM) of brain is affected by neuro-degeneration at initial stages within temporal lobe region and probability mapping is normally utilized as the input of classification problems. It also provides a quantitative analysis over spatial distribution of cells in brain with brightness of voxel that shows local GM quantity. Another technique of pre-processing is skull stripping. It is used over skull from images. Finally, motion correction has motion artifacts which are suppressed.

III. FEATURE EXTRACTION

Feature extraction techniques, based on neuro-imaging, data are used to develop quantified information set like shape, texture and volume of diverse brain parts. It will enable to identify disease pattern. The three phases of every classification problems are dimension reduction, feature extraction and classification. These steps are combined by deep learning models into one.

Noise in the raw brain scans comes from diverse sources and diverse levels based on scan type. Noise is owing to operator issues, neural activity, equipment and environment. Single brain scan comprises complex voxel patterns and large data quality, so it is very difficult to classify and interpret features.

Because of this reason, it is essential to extract small amount of distinct and pre defined representation areas indeed of complete brain image evaluation to classify images. Input data is managed due to its significance and it depends on extracted features. Different methods used for input data management are slice based, voxel based, patch based and ROI based. Voxel based method is used for voxel intensity measures taken from tissue components and different neuro-imaging modes (GM/WM in MRI).

In slice-based approach, [3] some properties are used to reduce 2D images, lowering sum of hyper-parameters. Various investigations use certain approaches to obtain 2 Dimensional portions from 3 Dimensional scans of brain. The central part of brain is taken by slice-based methods and ignores the rest. ROI methods [3] aim on specific brain parts are affected in earlier AD stages. This method is easily implemented in clinical practice. ROIs require knowledge regarding to abnormal areas and brain atlas like Automated Anatomical Labeling (AAL). Patch [4] is described as 3D cube. It identifies disease based patterns in brain with patch-based methods by extracting features from small image patches. Selecting maximum informative patches for acquiring image-level and patch-level features are the major challenge in patch-based methods. This method has been used in some of research for AD detection.

IV. OVERVIEW OF DEEP LEARNING MODELS FOR ALZHEIMER'S DISEASE PREDICTION

This section outlines some basic concepts of deep learning techniques and their architectures which might be determined in AD detection. It is possible to extract knowledge using Deep Learning (DL) from medical image as input data [3]. It helps radiologists to produce more accurate diagnosis. Two different deep learning methods are unsupervised and supervised. In neuro-image analysis unsupervised deep learning networks are used for feature extraction. The stage from ML to DL uses unsupervised DNN to extract higher level features. The unsupervised feature learning types are as follows. It is divided further into the following types.

A. Deep Belief Network

Deep Belief Network in short DBN [1],[2],[5],is a neural network made up of hidden layers and input layer. There exist connections among nodes of different layers. The nodes within each layer are not connected. RBM (Restricted Boltzmann Machine) is used as basic block of DBNs. It can be helpful for dimensionality reduction, regression and classification and feature learning. Important features can be extracted by RBM [6] and reconstruct input data. The basic architecture of RBM is shown in Fig. 1. Here ωij is weight among the nodes i & j. RBM architecture consists of different layers and nodes in one layer are connected with all other nodes in the next layer. Nodes at same layer are not connected.

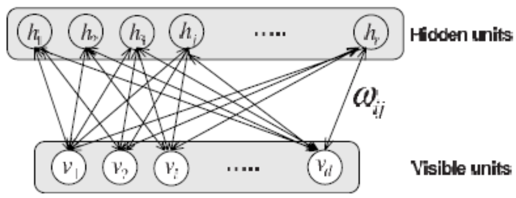

Fig 1. RBM Architecture

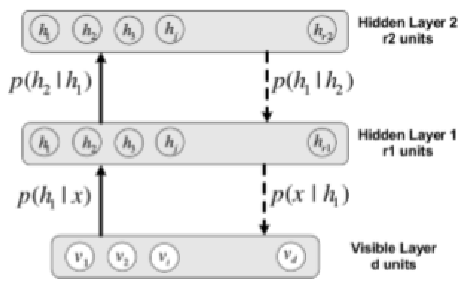

Fig2. Deep Belief Network

A fast unsupervised learning algorithm (Contrastive Divergence (CD)) [6] is used. The input of the hidden nodes is the inputs from visible layer multiplied with their respective weights. The bias is added with the sum of those products and then the activation function is and produces one output for each hidden node. This method modifies the weight of the connections by calculating the errors between training data and the reconstruction of input data. DBN is a stack of trained RBM. Each layer takes its input from last trained layer. Output taken from hidden layer is given to the visible layer of next RBM as an input and also uses the current biases and weights. RBM is trained in same process. This process is repeated till it meets out stopping condition. RBM is computationally very costly within the training process.

B. Auto Encoder

Auto encoder (AE) is a kind of ANN (Artificial Neural Network). It is an unsupervised learning method. Dimensionality reduction is possible by the use of an auto encoder, by training the network to avoid signal “noise”. AE [7] comes under NN including two phases: Encoder and Decoder. For classification, features learned in AE middle layer is extracted and utilized as pre-training phase during dimensionality reduction and feature extraction in unsupervised manner. Its structure is very simple and AEs power specification is extremely limited. A more enhanced version of AE called stacked auto encoder in which multiple AEs can be stacked. Stacked AE improves representational power using hidden unit values of AE as input to AE. Stacked AE has the capability to learn non-linear and complex patterns. Higher-level representations are learned along with the increase of the depth of the model. DL structures comprises of an AE and softmax layer. Several Types of auto encoders exist, they are Sparse, Denoising convolutional and variational auto encoders .For predicting Alzheimer's , it is possible to use AE an appropriate filter for convolution, and then any of the classification method can be used [7]. The architecture of auto-encoder is the Figure 3. Unsupervised models are mostly applied for extracting features. These outcomes are then classified by the help of a classification method or a classifier. Based on the above mentioned model, stacked Auto Encoder may drastically enhance non-linear and complex patterns. Advantages of AEs are to determine finest CNN parameter initialization.

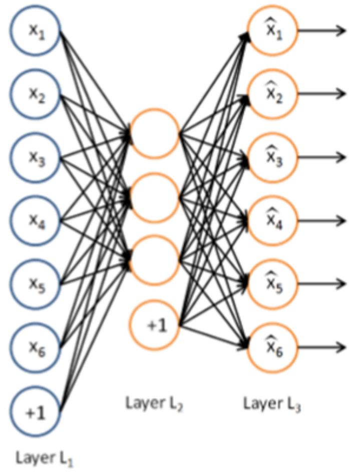

Fig 3. Auto-encoder

In supervised learning methods extracting the features and input classification are combined into a single model. Examples of such models are DNN, DPN, CNN, and RNN.

C. Deep Neural Network

Deep Neural Network [8] is an Artificial Neural Network and it comprises enormous hidden layers among input and output layers. DNN configuration is similar to conventional Multi-Layer Perceptron (MLP) network. It can discover complex correlations among input data and offers superior understanding towards data. MLP is based on supervised learning and it is used in various investigations for finding unknown conceptual patterns and corresponding correlations among them. Researcher cannot say that the method of training used in DNN is not good. The learning method also slow compared to SVMs.

D. Deep Polynomial Network (DPN)

Deep Polynomial Network is a supervised learning approach [9] based on DL. This method has better performance when compared with Deep Belief Networks and stacked Auto Encoder. We can increase representation performance of DPNs by stacking them into construct a deeper configuration. A stacked DPN (SDPN) [9], which is multi modal and comprises 2-stage SDPNs is anticipated for extracting the features in multi modal neuro-imaging data. The high level features that can be obtained from PET/MRI with the help of 2 SDPs. These features are given to SDPN to combine multi modal neuro-imaging data. The learned higher level properties include inherent features of every modality and the relations in between different models. The unsupervised method with SVM (linear) is used for training and classification.

E. Convolution Neural Network

CNN comes under DNN that is used for computer vision, which comprises of multiple layer neurons which are computationally connected. It is processed in a step by step manor and obtained enormous enhancement within computer vision area of research. CNN architecture [10] is included convolutional layers, pooling layers and fully connected layer. The CNN architecture is shown in the Figure 4 .

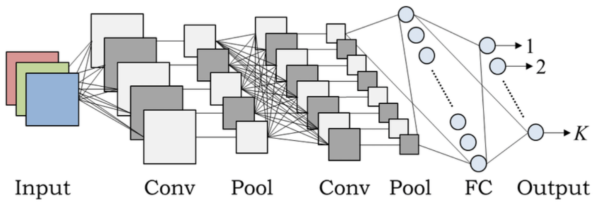

Fig 4. CNN architecture

The convolution layer has an important function to find discrete local lines, its edges and some other visual components. Filter operators have certain parameters and call it as convolutions, which are learned. By the help of kernel (i.e. small array of parameters which are learned),the pixel's local-neighbors are multiplied. The visual features are extracted by learning the kernels. A set of Filters are used to perform this process. Each filter used in this is square-shaped object that moves to image. Image values are added with filter weights.

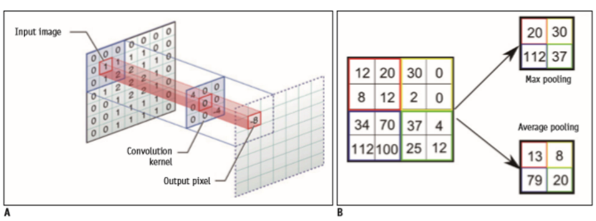

Fig. 5 Convolution method

The convolutional layer uses numerous filters and produces feature maps. Convolution is a baseline CNN component. It is utilized for segmentation and classification. To recognize large field view, mapping of features are diminished with pooling pixels together. By maximal propagation through average pooling, the convolutional layers eventually possess very less sensitive to deformations of target item during feature mapping. It is reduced by using pooling layer and remains robust to position and shape of images' semantic features. Max pooling is analytically applied. The above mentioned convolutional layers and the pooling layers are repeated several times. The Figure 5 represents convolution method and pooling methods. The fully connected layers [11] are merged with features from complete images and offer outcomes.

Deep CNN architecture includes various layers in NN, for that reason there are millions of weight parameters to compute. Therefore, it needs huge data samples for training and tuning. Generally, minimum data size requirement is based on radiologic image application. Moreover, there are some strategies to deal with its data size. Augmentation of data is one among them. By using those strategies approximately hundred cases/classes and offer appropriate outcomes. 2D CNNs are modeled to identify 2D image patterns. Various investigations have utilized 2D CNNs for 3D neuro-imaging. To build 3D, CNN [10] needs huge amount of parameters than 2D CNNs. There are 2, 3 and 5 convolutional layers can be used for image analysis.

3D images are obtained from neuro imaging. 3D CNNs are popular because of the spatial relationship among the images. 3D CNN may be a time consuming method and it is implemented using a small set of data to obtain a better architecture. Then the full size dataset is used for executing the final architecture. Keras [9] can be used to develop experiments including the final architecture. For detecting AD, It may suitable to use the whole image or some region of interest as the input. But this execution can also needed training huge amount of parameters over lesser dataset and results over fitting. 3D CNNs with 12, 11 and 5 convolutional layers were commonly used. Other one is a network with 7 convolutional layers, wherein three diverse filter sizes have been selected in 1st convolutional layer for seizing features.

One option is Cascaded CNNs. MRI and PET images were used for learned features. In this method, local images is converted to high-level features by using 3D CNNs. It consists of 4 convolutional layers, which are modeled on diverse local patches of image. Then, 2D CNN comprises 2 convolutional layers to merge related features and produces correlated image features. This can be merged with the last layers for AD classification. With supervised approaches, primary competition is among 3D [12] and 2D Convolutional Neural Networks (with/without Recurrent Neural Network).3D CNN performance is relatively higher than 2DCNN and capture 3dimensional details from 3D MRI scan. Training 2D CNN is easier to train but it is not efficient to encode 3D image's spatial information owing to kernel sharing absence across 3D.

Author [13] discussed about prediction of AD with deep 3D-CNN.This method has learned features from AD biomarkers and used over various datasets. 3D convolutional auto encoder also called 3D-CAE is used in 3D-CNN. The shape changes in sMRI (structural brain MRI) scannings is obtained by the help of 3D-CAE. 3 Dimensional CNN contains fully connected layers which are fine tuned for AD classification. 3D-CNN [14] uses MRI and PET to classify three binary class classifications (AD /. NC, MCI /NC, and sMCI / pMCI). Conventional unsupervised AE (auto encoder) is used to extract global features of three dimensional images. Huge parameters are evaluated using encoding and decoding layers. It is complex, expensive and needs huge data sets for training. For patterns extraction, which consists of large dimensional brain images and auto encoders were used for learning global features and it is fully connected among its layer's nodes. So stacked unsupervised convolutional auto encoder with convolutional weights and locally connected nodes to acquire local features from 3dimensional brain images along with voxel wise signal vectors was proposed [14] to overcome the above problem. Classification is done with 3D Adaptive CNN. Author [14] used the method CNN-LeNet which is used to classify functional MRI(fMRI)[15] data from normal controls, with accuracy of 96.85% over test data.

V. DATA SET

Dataset[16] is taken from Alzheimer s Disease Neuro-imaging Initiative [5][17] database (adni.loni.usc.edu) is used by most of the research paper in this area. The advancement of MCI and earlier AD by using PET, MRI and neuropsychological and clinical assessment is performed using ADNI dataset. Markers for earlier AD prediction are analyzed by researchers and clinicians to construct certain treatments and observe effectiveness with time reduction and clinical trial costs. ADNI database was developed to analyze AD's progression. It is done by gathering huge MRI data and PET images and blood biomarkers and cerebrospinal fluid investigations. Early AD stages can be diagnosed by the help of this database. Mainly three types of data are available; Data for healthy individuals, AD patients and those who suffered from MCI are available in ADNI database. OASIS attempted free distribution of MRI that includes two benchmark datasets. These include MRI data with 416 subjects (adults involving middle aged, young, suffering from dementia and not suffering from dementia) age ranges from 18-96. The dataset may include MRI data with 150 subjects with age range 60 to 96, which can be used for training the model.

VI. DEEP LEARNING PLATFORMS

Some of the deep learning platforms were studied, TensorFlow is one among them. It can be used with languages like Python, C++, and R to model DL models.

TensorFlow contains widely used two tools and they are:

1. TensorBoard :It can used for most effectual data visualization with network modeling

2. TensorFlow Serving: It is used for rapid deployment of recent methods with similar APIs and server architecture. It additionally offers merging various TensorFlow models that varies from traditional practices and extended to work with other data types.

Keras [9] is developed in Python and a high-level neural networks API. It has the competency to function over the top of Tensor Flow, CNTK, or Theano. It is possible to make fast experimentation using Keras. It gives a simple interface for quick prototyping purpose by making powerful NNs that functions with TensorFlow. It is significantly used for text generation, classification, summarization, tagging and translation and speech recognition. The Table I shows comparison of different DL approaches for AD prediction together with its accuracy. The accuracy of different methods over ADNI data set is listed with in the table. Most of the methods give good accuracy in test data.

TABLE I. Comparison of different Deep Learning methods

Methodology	Images	Dataset	Accuracy
CNN-LeNet	fMRI	ADNI	96.85%
3D-CNN	MRI and PET	ADNI	92.87%
Deep-CNN(DCNN)	MRI and PET	Publically available dataset	96%
Autoencoders and Softmax Classifier	MRI and PET	ADNI	87.76%
Sparse autoencoder	MRI	ADNI	AD vs HC=95.39% AD vs MCI= 86.84% HC vs MCI= 92.11%
Transfer learning on 2D CNN, LSTM network	MRI	ADNI	90.62
CaffeNet GoogleNet	MRI and PET	ADNI	CaffeNet:87.78% GoogleNet:83.23%

VII. CONCLUSION

AD is one of the most important reasons of loss of life in advanced nations. To detect AD early is completely hard, so application of computer-based systems, along with clinicians, has an important role in AD prediction. For performing this task, DL has given attention in current years. The evaluation with this investigation is about how DL (Deep Learning) could be used in developing AD detection systems. Deep Learning is used for decision making procedure to detect/ classify multiple stages of AD. This work started by defining AD and clinical features. Different biomarkers like MRI, PET, and fMRI were mentioned. Integrating the neuro-imaging modalities allows in AD prediction. Also, mentioned about the techniques of pre-processing, like normalization and registration. To deal with images, ROI and patch-based techniques had been suggested to be effectual in contradiction with slice and voxel techniques because of its capability to handle AD features in brain scan. Various deep learning methods are analyzed. With classification techniques, Convolutional Neural Network is frequently used with finest accuracy in contrary to other deep models. Comparison of different deep learning methods over neuro images along with its accuracy is also discussed.

REFERENCES

[1] Brian R Walker, Nicki R Colledge,Stuart H. Ralston, Ian D Penman: Davidson's Principles & Practice of Medicine,22nd Edition.

[2] Congling Wu, Shengwen Guo, Yanjia Hong, Benheng Xiao Yupeng Wu Qin Zhang, Discr- imination and conversion prediction of mild cognitive impairment using convolutional ne-ural networks :Quant Imaging Med Surg. 2018 Nov; 8(10): 992–1003.doi: 10.21037/ qims.2018. 10. 17

[3] Deep learning to detect Alzheimer's disease from Neuro-imaging: A systematic literature review. Computer

Methods and Programs in Biomedicine-2020.

[4] Pingu Liu, Shizhong Han,Zibo Meng,Yan Tong Facial Expression Recognition via a Boosted deep belief network: The IEEE Conference on Computer Vision and Pattern Recognition.2014DOI:10.1109/2014.233

[5] Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer's disease: International Journal of Neural system; Vol 26, No.7 (2016) 1650025

[6] Juan M.G´orriz and Javier Ram´ırez , Ense-mbles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer's Disease International Journal of Neural Systems Vol: 26 No 07, 1650025 (2016)

[7] Pushkar Batkoti,Manoranjan Paul, Early diag-nosis of Alzheimer's disease:A multi-class deep learning framework with modified k-sparse autoencoder classification,2016 , International Conference on Image and Vision Computing New Zealand(IVCNZ),IEEE DOI:10.1109 /IVCNZ.2016.7804459

[8] Siqi Liu, Sidong Liu, Student Member, IEEE, Weidong Cai1, Member, IEEE, Sonia Pujol2, Ron Kikinis,Dagan Feng, Fellow, IEEE, Early Diagnosis of Alzhiemer's Disease with Deep Learning,2014 , IEEE 11th International Symp-osium on Biomedical Imaging. Electronic ISBN: 978-1-4673-1961-4, DOI: 10.1109/ ISBI.2014.6868045

[9] Multimodal Neuro-imaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Dise- ase IEEE Journal of Biomedical and Health Infor- matics -DOI: 10.1109/ JBHI.2017.

[10] Karl Bäckström, Mahmood Nazari, Irene YuHua Gu, Asgeir Store Jakola , An Efficient 3D DeepConvolutional Network for Alzheim-er's Disease Diagonosis Using MRI Images. 2018 IEEE 15th International Symposium on Biomedical Imaging ISSN: 1945-8452 DOI: 10.1109 /ISBI. 2018.8363543

[11] June-Goo Lee PhD, Sanghoon Jun PhD,Young-Won Cho MS, Hyunna Lee PhD, Deep Learning in Medical Imaging. General Ove-review: .18.4.570

[12] Jyoti Islam , Yanqing Zhang :Deep Convo-lutional Neural Networks for Automated Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment Using 3D Brain MRI, International Conference on Brain Inform-atics BI 2018: Brain Informatics pp 359-369

[13] Adnan Qayyum , Syed Muhammad Anwar Muhammad Awais, Muhammad Majid: Medical image retrieval using deep Convolutional neural network.1016/j.neucom. 2017.05. 025

[14] Deep Learning Based Imaging Data Completion for Improved Brain Disease Diagnosis: Medical Image Computing and Computer Assisted Intervention –MICCAI pp 305-312

[15] Saman Sarraf Ghassem Tofighi, Classification of Alzheimer's Disease Using fMRI Data and Deep Learning Convolutional Neural Networks arXiv 1603.08631 2016. https://arxiv.org/abs/1603.08631.

[16] https://www.datacamp.com

[17] Yuming Hua Junhai Guo Hua Zhao. Deep Belief Networks and Deep Learning. IEEE DOI: 10.1109/ICAIOT.2015.7111524