EARLY DETECTION OF ALZHEIMER'S DISEASE USING DEEP LEARNING TECHNIQUES

Linta Oommen[1], Silpa Chandran[2], Vidhya L Prathapan[3], Krishnapriya P[4]

[1]Linta Oommen, APJ Abdul Kalam Technological University, Kerala. India.
[2]Silpa Chandran, APJ Abdul Kalam Technological University, Kerala, India.
[3]Vidhya L Prathapan, APJ Abdul Kalam Technological University, Kerala, India.
[4]Krishnapriya P, APJ Abdul Kalam Technological University, Kerala, India.
[5]Prof. Diliya M Khan, Assistant Professor, Department of Computer Science and Engineering, LBS Institute of Technology for Women, Kerala, India.

---***---

Abstract – *Alzheimer's is a neurological disorder. According to latest reports, it is the sixth leading cause of death. There are three stages of disease: Mild Alzheimer's, Moderate Alzheimer's and Severe Alzheimer's. Although the disease is incurable, earlier detection can help provide proper treatment hence prevent further brain tissue damage. The method used for Disease Detection is basically analysing the MRI Image. This may cause misdiagnosis and is very time consuming.Deep Learning is the method deployed to achieve the desired results and the model used is MobileNet where technique of transfer learning is applied.*

Key Words: Alzheimer's Disease Detection, Early Diagnosis, Deep Learning, Skull Stripping, MRI segmentation, Mobilenet

1. INTRODUCTION

Alzheimer's disease (AD) is a form of dementia. It is a type of progressive and neurodegenerative disease. The early diagnosis of the disease can help the disease from worsening even though it is incurable. Currently the method that is being currently utilized for Alzheimer's disease is analysing the magnetic resonance image (MRI) images , to arrive at a conclusion regarding the disease. Analysing MRI Images can help distinguish between an AD and NC due to its high resolution characteristics. However, the method is time-consuming, and since the symptoms vary from person to person it may lead to misdiagnosis. The main feature that, helps distinguish an AD from NC is the reduction in gray matter volume in the temporal lobe, hippocampal formation, insula and left thalamus of AD compared to NC.

The method that we are using for the disease detection is deep learning.Deep Learning portrays a higher accuracy while classifying. Deep learning automatically extract features. The two commonly used models for classification are VCG 16 Network Model and Mobile Net. The MobileNet network model was opened by Google on July 7, 2017 which is used in this project. The MobileNet network model uses a deep separable convolution unlike VCG which increases its computational power. This particular feature of Mobile net reduces complexity. It also caters to reducing redundancy.

Deep learning is an artificial intelligence method. Deep learning basically works like a human brain. Deep learning has the capability to learn from data that is unstructured or unlabelled. The artificial network used in deep learning has several levels. At the initial level certain information is gained by network. The information thus gained by the initial level is send to the higher level of the network. This continues as information gained at each level gets transcended to its next higher level and as it gets transcended a higher complex data is also evolved from the data obtained.

An example for identifying an object, the first level of a deep learning network identifies the edges and lines at a preliminary stage. This may be done by identifying the differently lighted areas. As mentioned before the information gained about the edges and lines in the first level gets transferred to the next higher level where further identification takes place. The next level puts together all the lines and edges identified to get an idea of the shapes present. The next higher level then uses the information generated by the previous lower level to get an abstract idea of what the object actually is with the avid features finally its in the final level that the object gets identified.

The deep learning has an higher complexity. The main problem that we face when it comes to using deep learning is the availability of data. A large dataset is required in order to obtain an accurate result. So in short only large dataset can provide us with best result while classifying a patient as AD, MCI and NC. In order to obtain the result accurately and to avoid any kind of misdiagnosis we made sure to include all kinds of images while training such as different ages, gender etc. The data or the images that we used are from the ADNI website. We noted that testing accuracy decreased when model was trained from scratch that is without using transfer learning. Hence in order to obtain the desired results with best available accuracy we resorted to use transfer learning method.

Therefore, transfer learning is adopted, which can be used for small datasets as well. Transfer learning is a technology that is used widely in computer vision and other fields.Transfer learning basically involves discovering and learning from information. The network layer gets initialized by using pretrained weights and dataset. After the network layers are initialized, the fully connected layer is transcended with the training data.Then the final layer of the fully connected layer is finetuned. The training speed was further increased by using pretrained weights that was provided by keras. Keras provided strong pretrained weights were used as initial values for better speed and accuracy

In this project, we used public brain MRI data from Alzheimer's Disease Neuroimaging Initiative (ADNI-http://adni.loni.usc.edu) Study. ADNI is an ongoing, multicenter cohort study. ADNI researchers collect, validate and utilize data, including MRI and PET images, genetics, cognitive tests, CSF and blood biomarkers as predictors of the disease. .Results are shared by ADNI through the USC Laboratory of Neuro Imaging Image and Data Archive (IDA)

The most common MRI sequences are:
1. T1-weighted
2. T2-weighted scans.

Repetition Time[TR] is the time between two successive pulse sequence applied at the same slice. The time between the delivery of RF pulse and receival of pulse signal is known as Time to Echo[TE].TE and TR Times are used for generating T1-Weighted Images. The T1 properties of an each tissue are what determines the characteristics and look of the Image. For Instance it is noted that the CSF[cerebrospinal fluid] is darker on T1-weighted imaging than on that of T2-weighted images. This evident feature helps distinguish between them. Fluid Attenuated Inversion Recovery (Flair) sequence resembles the T2-weighted image. The different in TE and TR length is what makes them different. For our project we have used the T1 weighted images from the ADNI dataset. The data used consist of patients belonging to three categories as CN, MCI, AD :

1.Control Health(CN) : The patient is not having Alzheimer's.

2. Mild cognitive impairment (MCI):Mild cognitive impairment (MCI) is the stage between the expected cognitive decline of normal aging and the more serious decline of dementia.

3. Alzheimer's disease (AD): This category patients are confirmed having Alzheimer's.

2. RELATED WORKS

[Akshay Iyer Et al, 2019] have proposed an idea to detect AD at an early stage using deep learning. Here the data is first skull stripped and then bias corrected using N4 Bias Correction and the registered to MNI 152 standard space which is known as affine registration. This process involves rotations, translations, etc. in order to align the scan with a standard brain structure Later on the segmentation process is performed to classify the tissues within the brain the final output after preprocessing is a 3D file which is then fed onto a 3D convolutional neural network. The features such as the grey matter, white matter and CSF is extracted and the fact that grey matter is less in the case of AD patients than in the normal control is taken into consideration for the classification purpose.

[Donghyeon kim Et al ,2018] Electroencephalogram (EEG) signal can be used to diagnosis Alzheimer's Disease (AD) at an Early Stage itself by obtaining a way to distinguish between healthy control (HC) and mild cognitive impairment (MCI). The metric used to distinguish HC and MCI is Relative Power[RP] which quantifies EEG pattern. The subjects in this paper comprises of two study groups. The first group consists of ten patients who were diagnosed as with mild cognitive impairment (MCI). The other group consisted of ten patients belonging to healthy control (HC) group. The subjects considered in this paper are diagnosed from the Chosun University Hospital (CUH, Gwangju, S.Korea) and Gwangju

Optimal Dementia Center located in Gwangju Senior Technology Center (Gwangju, S.Korea). The EEG recordings of all the subjects in this paper were acquired from 32 sites according to the 10-20 international system at a sampling rate of 500 Hz. The datset was divided into testing and training sets respectively. The feature extraction was done in the EEG with respect to RP. Three different frequency ranges were used. For classification a deep neural network of four layers was used. Then cross validation was done.

[Kanghan oh Et al, 2019] in the paper presents use of the end-to-end learning of a volumetric convolutional neural network (CNN) model inorder to classify four binary tasks as AD vs. normal control (NC) or progressive mild cognitive impairment (pMCI) vs. NC, stable mild cognitive impairment (sMCI) vs. NC and pMCI vs. sMCI) and then the outcomes are visualized based on CNN decisions. This requires no manual attention In the approach, they used a convolutional auto-encoder (CAE)-based unsupervised learning in order to get AD vs. NC classified, and supervised transfer learning was utilized for solving the pMCI vs. sMCI classification task .A gradient based visualization method was used for identifying relevant biomarksers. The contribution of the approach was validated by, conducting experiments with the help of data from ADNI and the results portrayed accuracy of about 86% for classification tasks which is good compared to other models. The temporal and parietal lobes were the regions that was used as crucial regions as far as classification was concerned.

[M.Latha Et al ,2019] in the paper proposes the idea of Roi to be inculcated to detect whether a person has the Alzheimer's disease or not .The grey matter and white matter of the brain were utilized for the same. The major steps involved are :Pre-processing ,Image Segmentation, ROI(Region of Interest),Classification .The images were taken from ADNI database. Image Pre-Processing involves grey scale conversion and noise removal using median filters. Segmentation of the image is done to highlight the areas of interest. The segmentation changes the image representation by processing label to each pixel of the input image. Roi is obtained from the segmented image by carrying out a binary mask operation on the segmented image. Prior to training feature extraction was carried which was done using wavelet transforms. Binary and multiclass classification was done, by taking into consideration the volumetric features and then this features are considered for the calculation of the amount of the grey matter distribution in each image. The dimensionality reduction was done using PCA. Several classifiers were used and the three of them SVM,IVM and RELM were found to be effective, Thus SVM was used for pattern recognition. The main idea behind using grey matter and white matter was that its was observed that the ratio of grey matter and white matter will be more for person having the diseases. So based on the ratio thus obtained it is concluded whether the person has Alzheimer's or not.

[Jyoti Islam. Et al, 2018] in the paper proposes a deep neural network for the Alzheimer's detection. The proposed network has several layers that perform each of their specific operations. The dataset used was OASIS, which has about 416 data samples. The layers are convolution, batch normalization, Relu and pooling. The SoftMax layer is used in the final output end with four classes as non demented, very mild, mild and moderate. The MRI images are not fed such to the proposed network, they are converted to several patches which are then fed into the proposed network. Sensitive training was incorporated to handle any imbalance in the dataset and a cost matrix is used at the final end, to modify or make any changes to the output. The weight is assigned based on the number of samples present in the class. The network actually utilizes only a small set of data. The data was divided into training and validation set. The validation was prepared using 10% of the training dataset. The performance was analysed and since this uses only a small set of data, it was found that this method can be improvised by incorporating more data. The model due to less data, suffered overfitting as well

[Siqi Liu.Et al, 2014] suggest a deep learning architecture which makes use of auto-encoders and also a softmax output layer. Early diagnosis is associated with the detection of MCI (Mild Cognitive Impairment).From MCI there is a high chance of transformation to AD or other forms of dementia. Two components are used to represent a learning structure of stack spaced auto-encoders and softmax regression layer. Auto encoders obtain deep representation of the original input. In sparse auto encoders consist of neural network with multiple hidden layers. Neurons of the input layer represent the original input vector. Each hidden layer represents the higher level representation of previous layer. The output layer is a sparse representation of input layer with same dimensionality as the input. The AD classification is perfomed by assing a softmax on the top of trained auto-encoder stack which contains hidden layers.

[Xiaoling Lu.Et al, 2019] proposed an idea to classify MRI Images using Mobile Net with the concept of transfer learning. It utilizes the pre-trained weights and MRI datasets form the initial network layer. Then the next full connected layer is simply retained after the previous step has been performed, and the final fully connected layer of the convolutional neural network is fine tuned to get better results and accuracy. The Keras provided pretrained weights which are highly efficient in terms of reducing training time and hence is made into use. Later on two hyper parameters such as width coefficient etc are also used.Thus the MRI images are first pre-processed, and then the features are extracted by the convolution layer,

the features thus extracted are then provided to the newly designed layer which is a fully connected layer that is then integrated to form a new network structure which uses transfer learning. The transfer learning method helps load the pre-training weights into the netwroks, and finally train networks and classify the MRI images.

[Dilek Mnasak.Et al, 2019] In this paper propses Random Forest as the method that can be used to eliminate some of the features which were not required. ADNI dataset was used for the studies. Data cleaning was performed as ADNI data included 12749 data from 1737 patients. The follow up of some of the patients up to 10 years was collected while some were collected only up to 2 years. This was repeated in every 6 months and some of data was found missing by them later. The study was conducted as solution at find the similarities when AD and MCI changes occurs. During data cleaning, healthy patients and patients without disease at present were deleted. Random Forest is used after data cleaning and Gini feature selection was used for taking the features into consideration. Using the features that were selected by Random Forest the Neural Networks were created.

[Marcia Hon.et al 2017]In the paper, transfer learning is used. The transfer learning uses architectures such as VGG and Inception are initialized with pre-trained weights from large benchmark datasets consisting of natural images, and the fully-connected layer is re-trained with only a small number of MRI images. They employed image entropy to select the most informative slices for training. Through experimentation on the OASIS MRI dataset, they showed that with training size almost 10 times smaller than the state-of-the-art, they reached comparable or even better performance than current deep-learning based methods.

[Suhuai Luo Et al, 2017] in the paper describes an automatic AD recognition algorithm that is based on deep learning on 3D brain MRI. The algorithm uses a convolutional neural network (CNN) to perform the AD detection. The three dimensional topology of brain is considered as one single unit for AD recognition in this paper hence the, was noted that the accuracy increased tremendously. The CNN used in this project or paper comprises of three consecutive groups of processing layers, two fully connected layers and a classification layer. In the proposed structure, each group comprises of three layers including a convolutional layer, a pooling layer and a normalization layer. The algorithm was trained and tested using the MRI data from Alzheimer's Disease Neuroimaging Initiative. The data used include the MRI scanning of about 47 AD patients and 34 normal controls. The experiment portrayed that the proposed algorithm delivered a high recognition accuracy of about 0.93.

[Jyoti Islam Et al, 2017]in the paper proposes a method for automated Alzheimer's disease early detection and treatment of the AD patients. A deep CNN model is being utilized for automated Alzheimer's disease detection and classification. The data used comes from theOASIS dataset. The model is something which was basically inspired by Inception-V4 network. Initially the preprocessing is performed and then, the input is passed through a stem layer. A stem layer comprises of multiple convolution layers, convolution layer, and a max pooling layer. Then the data that is obtained is undergoes passage through a series of inception and reduction modules. The input and output of all these modules pass through filter concatenation process. The softmax layer uses four different output classes which are nondemented, very mild, mild and moderate AD. The network takes an MRI image as input and extracts layer-wise feature representation from the first stem layer un till the last drop-out layer is reached. Based on this feature representation, the input MRI image is classified to any of the four output classes.

[Giovanni Montanna. Et al ,2015]in the paper proposes a a learning algorithm that accepts MRI images as input .The accepted images can be distinguished as healthy brains (HC) and diseased brains by the algorithm. A class of deep artificial neural networks and mainly the combination of sparse autoencoders and convolutional neural networks were used. The main aim of this approach was to identify how the 3D convolutions can be used in order to get better performance with an whole MRI Image.

[G.Stalin Babu Eet al, 2019] in the paper provides an overview of Prediction of Alzheimer's Disease using different machine learning methods. The major steps involved are Pre-processing, Feature Extraction and selection, Training and testing and Classification. The image Segmentation is the step involved in Pre-processing and it is performed using Fuzzy logic based segmentation algorithm. It is implemented using Fuzzy rule base and FIS targeted at detecting strong and weak edges of brain MRI images. Feature extraction involves extraction of specific features from the images .It is mainly used to decrease size of orginal data. Then SVM classifier is used for classification of disease.

[Kanganoh. Et al, 2020] in the paper mainly uses a Convolutional Autoencoder based unsupervised learning for AD and NC classification. The most important biomarkers are detected using gradient based visualization method which approximates the spatial influence of the CNN model. Autoencoders serve to provide dimensionality reduction as well.

Moreover the advantage is that it does not require labelled data. The input data are transformed into lower dimensional feature space during encoding phase and the data which has been encoded is then reconstructed to obtain the data in orginal space Fine tuning was performed for task specific binary applications and transfer learning was used for MCI classification.

[Soheil Esmaeilzadeh.Et al, 2019] proposes using a three-dimensional Convolutional Neural Network (3D-CNN)to be built using the Tensor flow framework. This model built could facilitate end-to-end classification. Apart from being able to provide the best classification performance when compared to other models it also helps identify biomarkers that are considered to be relevant. It was noted by the experimenters that the hippocampus region of the brain is very crucial when it come to the detection of the AD .An extensive hyper parameter tuning is made use of and a best architecture model is being made us of to the fullest. The MCI layer is fine-tuned a as well .In the long run they also found that the simple architectures provide better results compared to the other more complex architecture .As the chance of overfitting might be less.

[Garam Lee Et al, 2019] proposed a multi-modal deep learning approach to predict MCI to AD conversion using longitudinal cognitive performance and CSF biomarkers. Cross Sectional neuroimaging and demographic data was also used. Multiple GRUs were applied by them to make use of the longitudinal multi-domain data .The Results that they gained confirms that they achieved the better prediction accuracy of MCI to AD conversion only being used the longitudinal multi domain data. According to them this approach world be able to identify persons who are at the risk of developing AD and thus be able to provide adequate treatment.

[KR Kruthuka, Et al ,2019] proposes the need for proper selection of effective and better biomarkers (features) of brain MRI scans for AD, which would facilitate better prediction The multistage classification model for AD detection and image retrieval, were investigated in this paper. The swarm intelligent technique PSO for feature selection was the method employed by the experimenters and it was performed to get an adequate amount of information regarding the brain structural change, which is concerned with to the clinical detection of AD. The feature selection examines were: cortical thickness features, volume features, as well as a combination of thickness and volume. The multistage classifier which was utilized in this paper was found showing a good performance for AD detection when compared to other machine learning approaches and the image retrieval scheme followed by the method also portrayed appropriate results. The accuracy was high and performance was good.

3. SYSTEM FRAMEWORK

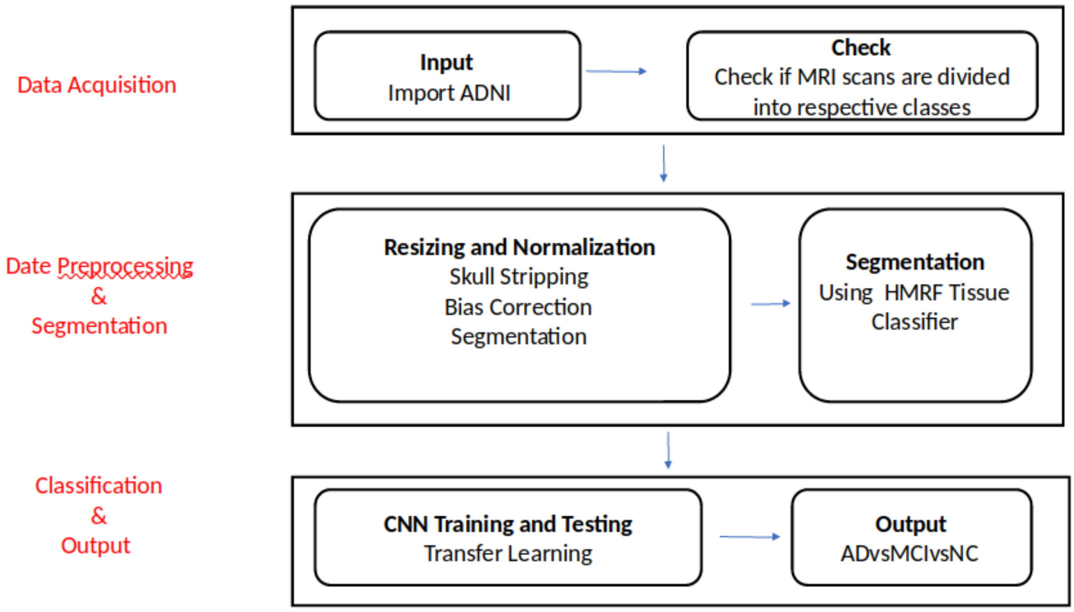


Fig- 1: System Architecture

As shown in the Fig 1, the system architecture composes Data Acquisition, Data Pre-processing, Segmentation and classification.

In data Acquisition that data is downloaded from ADNI website and divided into three folders, AD, CN, MCI. Using csv file it is verified whether the images are in the appropriate folders .The MRI images are first preprocessed, and then the bottleneck features are extracted by using the convolution layer part, and then the newly designed fully connected layer is connected to form a new network structure. We use transfer learning to load pre-training weights on new networks, and finally train networks and classify the MRI images. The pre-processing stage consists of Skull Stripping ,Bias Correction, Segmentation and Affine Registration. Skull Stripping removes the non-brain tissues. Bias Correction ensures that bias field is appropriate. Segmentation partitions the images based on different regions and affine registration standardizes the images.

3.1 Skull Stripping

Skull stripping is the process of removing the non-brain tissues from the brain tissues. It is performed using the Deep brain Extractor available in the deep brain library. The dataset contains the files with .nii extension. nii files are files that contain Nifti.1 Data format by Neuro Imaging Information Technology. The nifty file consists of two parts the header file and image file. The header file contains meta data ie the data about the image array ,details about dimension, dimensions present in the affine data, spatial arrangements etc. The mathematical morphology is used for skull stripping. The images to be skull stripped is first of all converted to a binary Image, using Otsu threshold method. The algorithm consists of the images as two classes which separates the two classes so that their intra class variance is minimum and their inter-class variance is maximal. The two components brain and the skull are differentiated using the erosion property. The pixel values 1 that is relevant areas that is brain part is retained while the parts containing pixel 0 is suppressed. SoftMax was utilized to generate output and cross entropy .The result obtained is compared with ground truth tables which uses weighted entropy as well. The output thus obtained is compared with ground truth tables gives us weighted entropy. First k pairs of training image*(X)* and labels *(Y)* as $\{X^i, Y^J\}$j=1,2..k is assigned. After which the result obtained is then flattened into 1d vectors where i and j .This result obtained is the voxel of the image order .Softmax classifier is used to calculate the posterior probability of voxel I with label l:

$$\mathrm{p}(y_i=\mathrm{I}|\mathrm{X}(\mathrm{M}i)) = \frac{e^{\mathrm{f}y_1}(M_i)}{\sum_{k'=1}^{k} e^{\mathrm{f}y_1}(M_i)} \quad (1)$$

The weighted cross entropy is calculated as:

$$\mathrm{Loss}=-\sum i \log\left(\mathrm{p}(y_{\mathrm{i}=}\mathrm{groundtruth}|\mathrm{X}(\mathrm{M}i))\right) \quad (2)$$

3.2 Bias Correction

Bias field signal is a low-frequency and very smooth signal that corrupts MRI images. A preprocessing step is needed to correct for the bias field signal before submitting MRI images to the algorithms. What we have used here is the N4 Bias Correction function from SimpleITK library.

The N4 bias field correction algorithm is a popular method for correcting low frequency intensity non-uniformity present in MRI image data known as a bias or gain field. The method has also been successfully applied as flat-field correction in microscopy data. This method uses a simple parametric model .The advantage is that it does not require tissue classification. ITK(Insight Segmentation and Registration Toolkit) is an open-source platform that gives developers an extensive suite of software tools that can be used for analysis of Images. SimpleITK is also a simplified layer that is built on top of ITK, which is done in oreder to incorporate its effectiveness in rapid prototyping, education, interpreted languages.

The characteristics of SimpleITK are as follows:
1. C++ library
2. Object-oriented
3. Provides a simplified, easy-to-use, procedural interface without templates
4. Is distributed under an open source Apache 2.0 License.
5. Binary distributions for Python and Java

High-dimensional mappings are computed by ANT to capture the brain structure and its corresponding function. It allows users to organize, visualize and statistically explore large sets of medical images. It combines imaging modalities and information that are linked to it with respect to space and time, and works across species or organ systems with minimal customization. ANTs relies on the Insight ToolKit (ITK) and to ANTs developers contribute to it. ANTs is popularly considered a state-of-the-art medical image registration and segmentation toolkit. To segment an MRI Image we minimize using following cost function:

$$\mathrm{J(U,V)}=\sum_{\mathrm{i}=1}^{N}\sum_{\mathrm{j}=1}^{c} u_{\mathrm{ik}}^{m}\ d^2\ (v_i, x_k) \quad (3)$$

Data points X={$x_1, x_2 \ldots\ldots x_N$} represents MRI pixels,U={$u_1, u_2 \ldots\ldots u_N$}, are the membership values.The constraints used for optimizing are:

$$f(x) = \begin{cases} \sum_{\mathrm{k}=1}^{c} u_{\mathrm{ik}} ,>0;\mathrm{i}=\{1 \ldots, \mathrm{c}\} \\ \sum_{\mathrm{k}=1}^{c} u_{\mathrm{ik}} ,>0;\mathrm{k}=\{1 \ldots, \mathrm{N}\} \end{cases} \quad (4)$$

Mathematical model of Biased MRI image is

$$y_k = x_k B_k \quad (5)$$

y_k,x_k,B_k are the corrupted image ,orginal image and bias signal field. We substitute 5 in 3:

$$\mathrm{J(U,V)}=\sum_{\mathrm{i}=1}^{N} \sum_{\mathrm{j}=1}^{c} u_{\mathrm{ik}}^{m} \left\| \frac{y_k}{B_k} - v_i \right\|^2 \quad (6)$$

3.3 Segmentation

Image segmentation may be defined as a technique, which partitions a given image into a finite number of nonoverlapping regions with respect to some characteristics, such as grey value, texture distribution, etc. It is represented as a collection of contours or masks where each segment is assigned a unique grayscale value or colour. It can be defined as a compilation of methods that allows to interpret spatially close parts of an image which can then be identified as an object. Spatial closeness natural qualify as Regions and they form the fundamental principle of segmentation. The object is everything what is of interest in the image. The rest of the image is background. The approach is similar to that used in pattern recognition, i.e., division of the image into set of equivalence classes. Segments can be used to separate tissue classes. A T1 weighted structural image is segmented by using Bayesian formulation. The observation model is termed as Gaussian distribution or Markov Random Field(MRF) is used to model the priori probability of the context-dependent tissues of the brain .This is based on three tissue probability. In this case MRI image elements are typically classified into three main tissue types, white matter(WM), grey matter(GM) and cerebrospinal fluid (CSF).Segmentation of the image is done using Hidden Markov Random field model (HMRF) method ,which is a stochastic process generated by Markov Random field whose state sequence cannot be observed directly ,but which can be observed directly but can be estimated through observations. The result of segmentation is a set of objects that can be identified and quantified individually .This type of 3d image volumes requires pre-processing ,which is carried out initially. K means clustering is the underlying algorithm that is being utilized to separate into the three classes.

Each observation is a l*l transformation matrix, where l denotes the number of states. The tissue classifier HMRF is utilized for the segmentation purpose. The tissue classifier class contains the methods for tissue classification using Markov Random Fields modelling approach. This method uses the maximum a poster ion-Markov Random Filed approach for segmentation by using iterative conditional modes and Expectation Maximisation to estimate parameters.

Expectation Maximisation is used for accurate segmentation and to get the optimal solution. Beta float is used as smoothing parameter which when increased in number increases the smoothness of the output. The main advantage of the HMRF model derives from the way in which the spatial information is encoded through the mutual influences of neighbouring sites. After the segmentation is completed, a matrix is obtained.

In order to preserve the points, straight and parallel lines affine transformations are used. Affine transformation uses a angle of rotation that is clockwise to the typical geometry unit circle of angles being measured in counter clockwise with 0 starting from positive x axis and hence negative of angle is often applied. Consider a point *X* in affine space, then every affine transformation on *X* can be viewed as the composition of a linear transformation on *X* and a translation of *X*. An affine transformation does not preserve the origin of the affine space like linear transformation. Hence, every linear transformation turns out affine, but not every affine transformation is linear .Iteration conditional modes are used to arrive at optimal solution. The iterational conditional modes is a deterministic algorithm used for obtaining local maximum of the joint probability of a Markov Random Field. The fuzzy c-means algorithm is based on minimizing the objective function which is the weighted sum of squared error of the clusters:

$$\mathrm{J(U,V)}=\sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^{m}\ d^2\ (x_i ,V_j),\mathbf{U}=\{U_{ij}\}, \mathbf{V}=\{V_j\} \sum_{j=1}^{c} U_{ij}=1,0 \leq U_{ij} \leq 1 ,\mathrm{i}=1,2 \ldots\ldots \mathrm{N} \quad (7)$$

Where X is is the data set, N the number of data items, c is the number of clusters with 2 ≤c≤ N, U_{ij} is the degree of membership of xi belong to the jth cluster, m is the weighting exponent on each fuzzy membership, V_j is the prototype of the centre of cluster j and d2ðxi; vjÞ is the distance between data xi and cluster center V_j.

Lagrange Multiplier is used and constrain in Equation (7), the derivatives of U_j and V_j in J(U,V) can be found:

$$U_{IJ} = 1/\sum_{k=1}^{c} \left(\frac{(x_i - y_j)}{(x_i - y_j)}\right) \text{^(2/m-1)} \quad (8)$$

$$V_{IJ} = \frac{\sum_{i=1}^{N} u_{ij}^{m} X_i}{\sum_{i=1}^{N} u_{ij}^{m}} \quad (9)$$

3.4 Mobile Net

The convolutional neural network that we used in our project is Mobilenet. The main reason for using Mobilenet is due to its lightweight architecture. Mobilenet uses depthwise separable convolutions that basically means it performs a single convolution on each colour channel rather than combining all three and flattening it. Deep convolution is more compared to standard convolution. The drawback is that it just filters the input channels and does not integrate it to create new features. Hence an additional layer requirement for the calculation arises. The depth-wise separable convolutions makes the neural networks light weight in architecture. The Depth-wise separable convolutions are divided into 2 operations known as Depth-wise convolutions and Point-wise convolutions. In depth-wise convolution operation the convolution is applied to a single channel at a time rather than in the case of standard Convolutional neural networks in which it is done for all the channels. Next in the pointwise operation, a 1×1 convolution operation is applied on all the channels.

Due to these two operations done separately the depth wise separable convolution network performs much lesser multiplications as compared to a standard convolutional neural network where all the process are done in one step.

Mobile Net is a linear stack of multiple neural networks. Initially the flat layer is created. The flat layer then transforms the Input from the previous layer.Then a hidden layer is created which consists of 256 neurons and use the SoftMax activation function. Then, a Dropout layer is added which is done to avoid overfitting. The convolution of the MobileNet network is separable convolution .Standard convolution filters features. The features are filtered based on convolution kernels and feature combinations and thus new representations are created. The filtering and other steps are then combined into to form two new steps by making use of deep separable convolution. This particular steps performed previously helps to enormously reduce computational cost. The deep separable convolution mainly serves to help break the interactions between size of kernel and outputs.

The MobileNet network model makes use of the depth-separable convolution mainly to reduce the computational complexity. The Mobile Net network model adds two hyperparameters: the width coefficient and the resolution coefficient. Due to small size it is used with many applications. Width coefficient is the parameter used for computations. It is simply the ration between total number of convolution kernels that can be used by each module to the standard Mobile Net. It reduces computational cost as well as the need for secondary parameters.

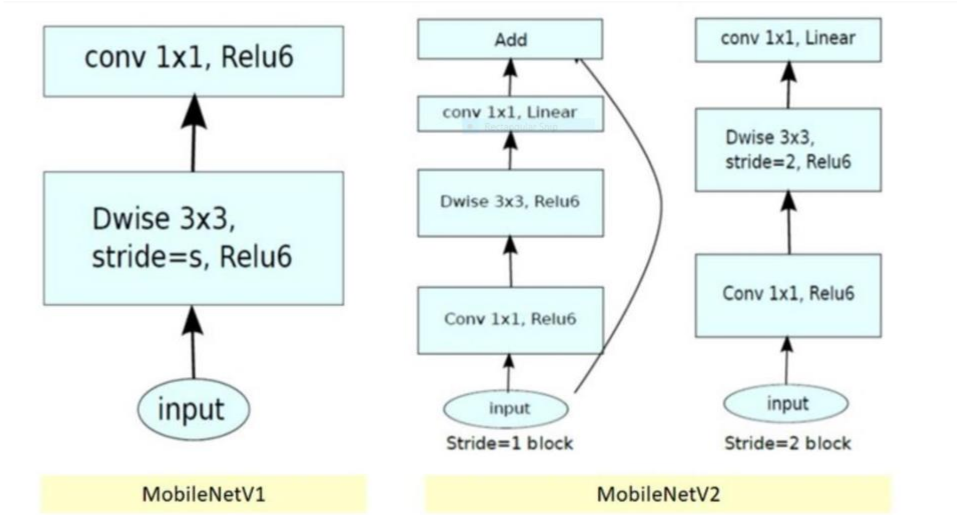


Fig-2: MobileNet Architecture (Sik-Ho Tsang 2019)

Transfer Learning is used for a more accurate result. Transfer Learning is the process of storing knowledge gained by solving a problem and then applying it to different but related problems. The reusing and transferring of previously learned tasks for the learning new tasks has the potential to significantly improve the sample efficiency of reinforcement learning. Therefore, the previous knowledge of the model acts as an input for the model which in turn helps in increasing it's accuracy. It involves the idea of freezing the base layers and adding new layers as per the requirement of the problem to be solved. In this experiment additional dense layer and a softmax layer is added and the loss method used is categorical cross entropy.

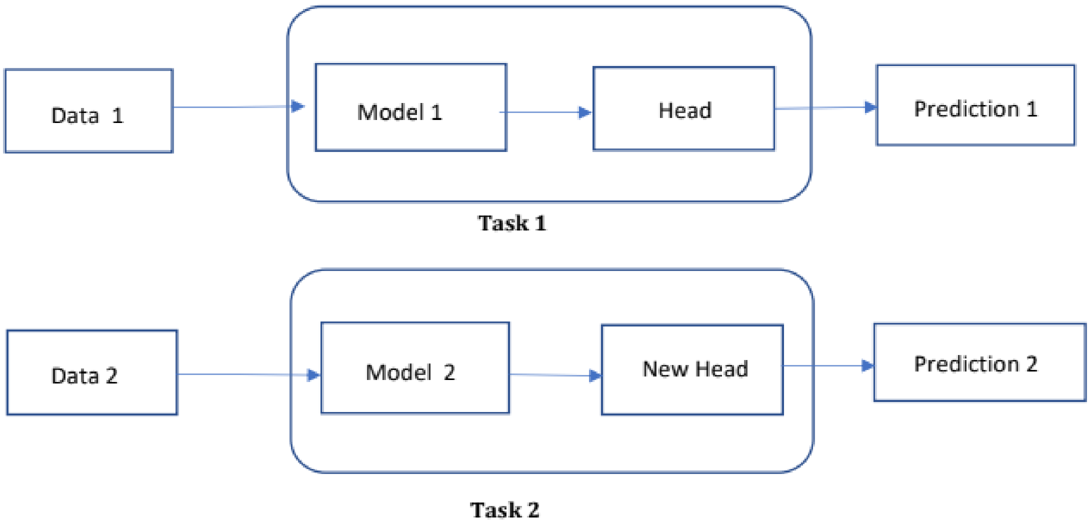


Fig - 3-: Transfer Learning

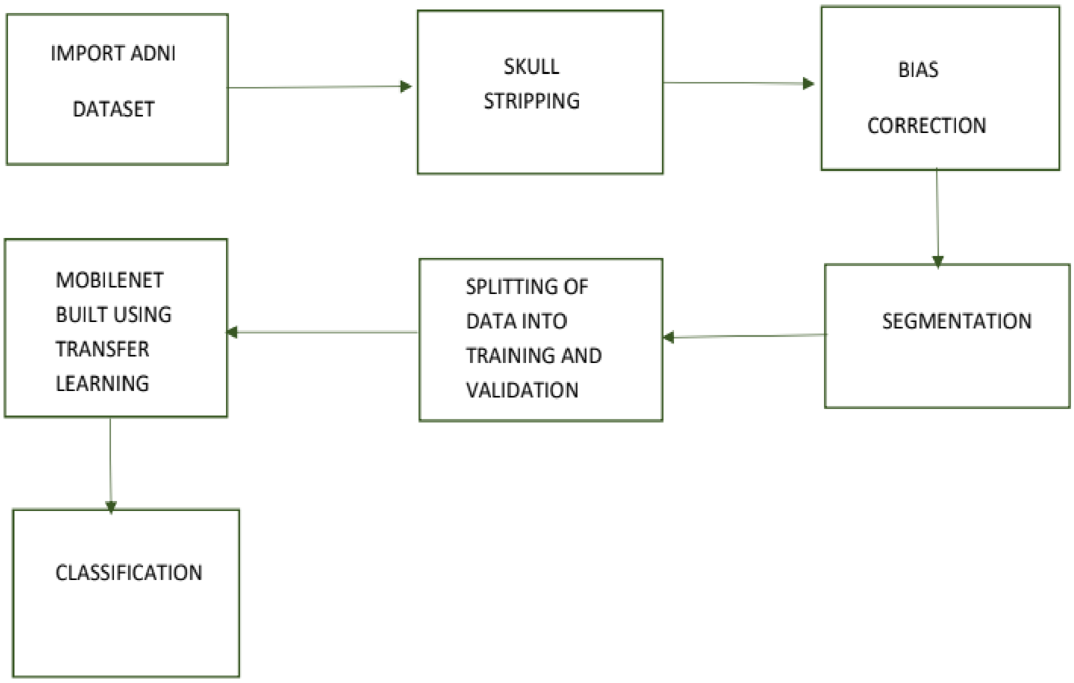


Fig - 4: Data flow Diagram

Initially the dataset is imported from the ADNI website and they are imported .The images thus downloaded are then skullstripped to remove the non- brain tissues as only the brain tissues are considered for the detection of the disease, after which the skull stripped images are Bias Corrected to a particular signal and then segmented into the three classes gm,wm and csf. The segmented images are then changed to MNI space to bring uniformity among all the images. The data thus obtained into training as well as the test data .The Training set is used for the training of the Mobile net which is the network that is being used.The network is built using transfer learning .

4. PROPOSED METHODOLOGY

Step 1: Import the ADNI Dataset D

Step 2: Preprocessing of images is carried out

Step 2.1: Skull Stripping: The Brains are skull stripped using Otsu Threshold by generating a mask.

Step 2.2: Bias Correction: Bias Field is calculated using Modified Fuzzy C Means Algorithm and Non Uniformity is removed.

Step 2.3: Then the images thus obtained are segmented into three classes grey matter, White Matter and CSF using tissueHMRFClassifier which uses k means Algorithm

Step 3: The Mobile Net model is built using Transfer Learning.

Step 4: Input the pre-processed dataset D into Mobile Net Model.

Step 5: The images are then classified as AD, MCI or NC

5. EXPERIMENT METHODOLOGY

5.1 Experimental data

The MRI scans of 1,350 patients were acquired from the ADNI website (http://adni.loni.usc.edu/). Here, first of all the 3D images in .nii extension are downloaded and pre-processed by performing skull stripping, bias correction and then segmentation. The segmentation results in segmenting the brain into three classes which is grey matter, white matter and cerebrospinal fluid. After the segmentation, three 2D image slices were obtained from the segmented brain using matplotlib which gave the axial, coronal and sagittal views. These images were then augmented to increase the size.The transformations involved a zoom range of 0.05, width shift range of 0.05,height shift range of 0.05, shear range of 0.05 and horizontal flip. In this manner, the 2D image dataset was formed for each of three classes AD,CN and MCI. The dataset contained a total of 40,000 images in each class where each image had a dimension of 256x256 and a format of png. The images were processed in grayscale mode. Using this dataset the MobileNet was trained with the technique of transfer learning. A learning rate of 0.01 and an Adam optimizer was made into use along with a softmax activation layer and categorical cross entropy. The experiment was conducted entirely on Google Colaboratory. Also, the deep learning framework of Keras was used with Tensorflow as the backend.

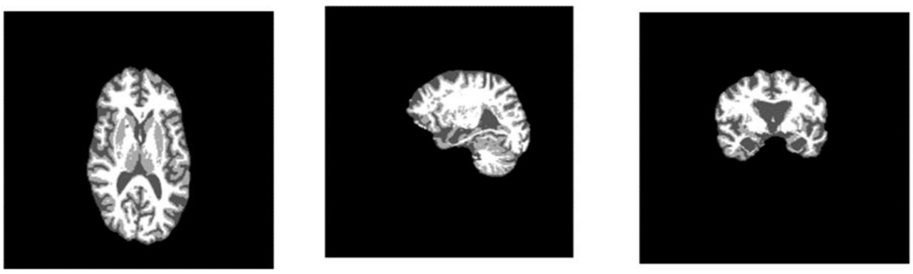
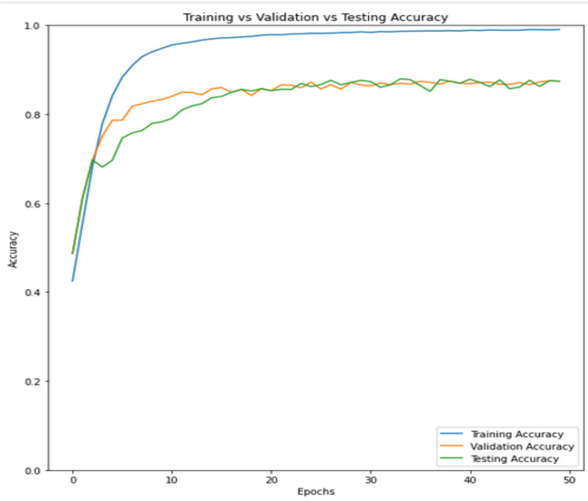


Fig -5: Axial (left), Saggital (centre), Coronnal(right)

5.2 Analysis of results

Here the results of the experiment conducted on the MobileNet Model is provided. The goal is to analyze how well the model is able to identify between AD, CN and MCI where CN refers to the normal patients.

5.3 Graphs



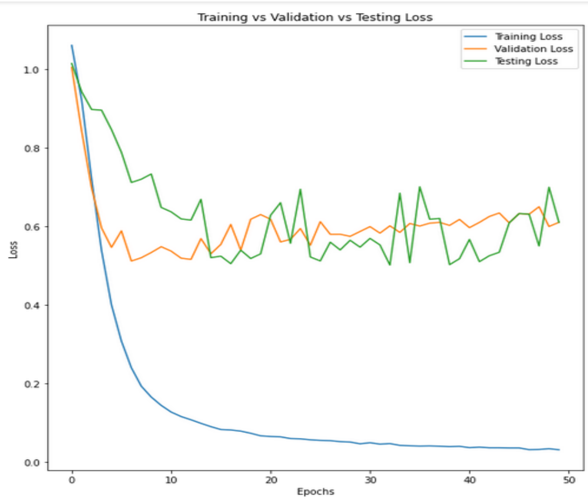


Fig-6: Accuracy vs Epochs in Training, Testing, Validation Data (Left) and Loss vs Epochs in Training, Testing, Validation Data (Right)

Figure 6 shows the training, testing and validation accuracy of the MobileNet Model plotted against 50 epochs. It is observed that as the training time increases, the model's validation accuracy also increases. When the iterations reach 21 , the model already attains an accuracy of 85% on the validation set and an accuracy of 97% on the training set. It can also be seen that as the number of epochs increases, the validation loss decreases.

6. CONCLUSION

The experiment has been conducted using the technique of deep learning. Here the chosen model is MobileNet due to its lightweight architecture and high performance. The method of transfer learning has also been applied in order to reduce the computational cost. The results of the experiment can be summarized as follows : The MobileNet Model can be used for detecting the AD and MCI patients from normal patients as it acquires an accuracy of 85%. Then additional improvements on accuracy can be made by increasing the size of the dataset. Then further, more features can be taken into consideration other than the grey matter, white matter and cerebrospinal fluid in order to develop a more accurate classifier model.

REFERENCES

[1] Gerardin, E., Chételat, G., Chupin, M., et al. (2009) Multidimensional classification of hippocampal shape features discriminates Alzheimer 's disease and mild cognitive impairment from normalaging. Neuroimage, 47(4): 1476-1486.
[2] Yang, J., Pan, P. L., Song, W., et al. (2012) Voxelwise meta-analysis of gray matter anomalies in alzheimer's disease and mild cognitive impairment using anatomic likelihood estimation. Journal of the Neurological Sciences, 316(1-2)
[3] Falahati, F., Westman, E., Simmons, A., (2014) Multivariate data analysis and machine l learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. J Alzheimers Disease 41(3), 685–708
[4] Salvatore, C., Battista, P., Castiglioni, I., (2016) Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines. Curr Alzheimer Res, 13(5): 509.
[5] Yang, C.H., Yu, C.J., (2017) Alzheimer's disease classification algorithm based on fusion principal component analysis and linear discriminant analysis. J Xiamen Univ NatS ci, 56(2): 226–230.
[6] Zhou, W., Wang, Y., Xiao, H.B., Cao, L.H., (2018) Auxiliary diagnosis of Alzheimer's disease based on KPCA algorithm. Chinese Journal of Medical Physics, 35(4): 404–409.
[7] Vaithinathan, K., (2019) A Novel Texture Extraction Technique with T1 Weighted MRI for the Classification of Alzheimer's Disease. Journal of Neuroscience Methods, 318: 84–99.
[8] Uhlich, M.; Greiner, R.; Hoehn, B.; Woghiren, M.; Diaz, I.; Ivanova, T.; Murtha, A. Improved Brain Tumor Segmentation via Registration-Based Brain Extraction. Forecasting 2018, 1, 59–69.
[9] Tosun, D.; Rettmann, M.E.; Naiman, D.Q.; Resnick, S.M.; Kraut, M.A.; Prince, J.L. Cortical reconstruction using implicit surface evolution: Accuracy and precision analysis. Neuroimage 2006, 29, 838–852.Appl. Sci. 2019, 9, 569 13 of 15
[10] Kalkers, N.F.; Ameziane, N.; Bot, J.C.J.; Minneboo, A.; Polman, C.H.; Barkhof, F. Longitudinal brain volume measurement in multiple sclerosis—Rate of brain atrophy is independent of the disease subtype. Arch. Neurol