Alzheimer's Disease Classification Using Deep Learning Techniques

Abhishek Nagrecha
Department of Computer Science
Lakehead University
Thunder Bay, Canada
nagrechaa@lakeheadu.ca

Rapelli Sai Rohit
Department of Computer Science
Lakehead University
Thunder Bay, Canada
srapelli@lakeheadu.ca

***Abstract*— Alzheimer's Disease (AD) is the most severe type of brain disorder found mainly in people 60 years of age or over.The latest developments on Multimodal Neuroimaging(MN) data have allowed the identification of the disease in life that was a breakthrough in neurosciences. Early diagnosis of AD is essential for the progress of more prevailing treatments.Recently, substantial focus has been paid to applying Deep Learning(DL) to early identification and automatic diagnosis of AD as rapid progress in neuroimaging techniques has generated large-scale MN data.To save the overall time and energy of the doctors and increase their effectiveness in saving the lives of the sufferers, we suggest a classification of the AD using deep neural networks such as CNN,ANN and MLP.This research is aimed to provide an overview and critical assessment for the early identification of the AD to help the physicians in providing appropriate treatment to the patients based on the four different classes which we have considered in our dataset.**

I. Introduction

Alzheimer's disease (AD), a form of dementia, being characterized by gradual thought and behavioural difficulties that occur in the middle or old age. It has been found that Alzheimer's disease accounts for about 60-80 percent of cases of dementia and, on median, a person with Alzheimer's existence is four to eight years after diagnosis, but may survive for up to 20 years, based on certain circumstances. The neurological characteristics may include the presence of neuritic plaques in the cortex, and overt brain cell calcification. The signs usually grow gradually and are severe enough to interfere with everyday activities. While oldness is the primary risk factor, AD is not merely an old-age disorder. During the early stages, the lack of memory is moderate and in the latter stages, interaction with the patient and their ability to respond degrades dramatically. Current therapies cannot delay the progression of Alzheimer's disease (AD), but early intervention will help reduce the progress of the condition and can increase the quality of life for patients. The early signs of Alzheimer's most often found include difficulties in retrieving recently acquired knowledge. Our minds change much as the rest of our bodies as we mature. All of us will experience some sluggish thinking and occasionally challenges reminding us of other things. However, significant lack of comprehension, uncertainty and other big shifts in the way our brains operate can be an indication that the neural connections have started to malfunction.

The first manifestation of Alzheimer's disease as stated before is difficulties in processing recently discovered knowledge because Alzheimer's changes typically begin in the part of the brain that affects learning. When Alzheimer progresses through the cortex, it leads to more debilitating symptoms, including improvements in claustrophobia, mood, and behaviour; growing uncertainty about activities, time , and place; irrational assumptions about family , friends, and caregivers; quite extreme cognitive impairment and behavioural changes; and difficulty talking, chewing, and standing. It's real that doctors can't provide a treatment if you have Alzheimer's syndrome or the underlying disorder. Yet prenatal recognition can be helpful.

Learning what to do is just as important as knowing what to do. When an individual has a treatable disorder that causes cognitive dysfunction or complicates the disability in any way, then specialists can start care. Doctors mostly prescribe opioid and non drug treatments for people with Alzheimer's dementia to help reduce the burden of the condition. Researchers also recommend medications which help speed down cognitive impairment and other cognitive abilities,hence finding a professional as quickly as possible is of paramount significance.

A. Detection of the Alzheimer's Disease

In order to detect Alzheimer's disease most of the hospitals and clinicians ask for the brain-imaging studies for them to create pictures of our brain
1] Magnetic resonance imaging (MRI)- MRIs may help to accumulate important markers, such as swelling, bleeding and structural problems.
2] Computed tomography (CT)- CT scans use X-ray pictures that will help the doctor look for irregular neural traits.
3] Positron emission tomography (PET) scan- A PET test utilizes a radioactive agent to identify contaminants in the body known as a tracer. PET scans are of various types. The far more commonly employed PET scan is a PET scan for fluoro-deoxy glucose (FDG), which can classify brain regions with a reduced metabolism of carbohydrates.

The pattern of differences in metabolism will differentiate between various forms of degenerative disease.Such scans were recently designed to identify amyloid protein clusters (plaques) correlated with Alzheimer s disease, however this type of PET scan is usually used in the research settings.A specialist could use brain mri devices under some cases to figure out whether the patient has elevated levels of beta-amyloid, a symptom of Alzheimer's; normal levels would indicate that Alzheimer's is not really the cause of dementia. Psychologists are researching other brain MRI methods so that they can detect and monitor Alzheimer's development further. 2) Cognitive tests performed by a specialist A psychiatrist or other expert, such as a forensic psychologist, can administer such pen and paper assessments, such as the Montreal Cognitive Assessment (MoCA), which evaluates interpretation, preparation, problem-solving, logic, and recollection. Will take time to make the diagnosis.

The diagnosis may be made by a family doctor or a doctor. The specialist may, or may not, refer patients to other healthcare providers before making the diagnosis. A physician, psychiatrist, haematologist, geriatrics, nurse, counsellor or occupational therapy may be interested in this.

In this research project, our aim to discover whether a person is suffering from Alzheimers or not.To be more precise, we want to predict the person's exact state of mind, so we can perfectly categorize his/her condition. For that, we will make use of popular Deep learning techniques to prepare our desired data for classifying the users and making predictions according to the result of a classifier, when applied on a dataset containing Persons details. Our objective is to analyse the performance change for multiple deep learning models with different layers when applied to the original and pre-processed dataset. Our motivation comes from the fact that we are very sure that the results of these models will help the doctors and hospitals in the future to provide the appropriate medications and treatment to a patient according to the category to which one may belong and that will help the patients to cope up with Alzheimer in a highly effective manner. This will also increase his/her chances of overall recovery and will make a lot more difference in their lives. For the hospitals, this will help them to disperse their resources in the correct directions and speed up the process of recovery.

We decided to work with Magnetic Resonance Imaging (MRI) for this purpose, which is typically used to examine the structural brain regions because of its fine spectral resolution and soft tissue contrasting ability. MRI is commonly considered to be associated with less safety risks compared with other modalities such as computed tomography (CT) and positron emission tomography (PET) [1] [5].

During the last two decades there has been enormous success in treating brain trauma and studying brain structure with MRI[2]. Disorders such as Alzheimer's disease (AD), and brain-related multiple sclerosis[3] can be detected utilising MRI. An atrophy of the tissue is a common predictor used to diagnose AD. Brain MRI segmentation, taken at various times, is often used to calculate structural changes in the brain.

II. Literature Review

CNNs have demonstrated excellent results in category, classification, identification, and retrieval of images as a best learning algorithm for understanding the quality of images. CNN has the potential to use spatial or temporal data correlation as its attractive attribute. The CNN topology is divided into several learning steps, consisting of a mix of convolutional layers, non-linear processing units and sampling layers. While dealing with complex learning issues, deep architectures also have an advantage over low-level architectures [14]. The layer-specific stacking of several linear or non-linear processing units helps one to understand complex representations at different abstract levels.

As a result, deep CNNs have shown a major improvement in performance over traditional vision-based models in Alzheimer's disease diagnosis consisting of hundreds of image categories [15]. The finding that the deep architectural capabilities of a CNN will boost the use of CNN in image classification and dividing tasks. The recent success of deep CNNs was also mainly attributed to the availability of big data and hardware advances. Empirical work has shown that deep CNNs can learn the invariant diagnosis depictions and achieve human performance when appropriate training data is provided. In order to extract meaningful representation from a wide range of unlabelled data, deep CNNs can also be used as a supervised learning tool[16]. The use of a transfer learning principle (TL), has recently shown that various levels of features, including both low and high levels, can be transferred to a generic recognition function. Various changes in the CNN learning methodology and architecture were made between the late 1990s and 2000 to make CNN scalable to big, heterogeneous, complex, and multi-class problems. CNN technologies cover different aspects such as processing unit adjustment, optimisation techniques for parameters and hyperparameters, design patterns and layer connectivity, etc. After AlexNet's outstanding achievement in the ImageNet dataset in 2012, CNN-based applications became prevalent [17]. Since then, major developments have been proposed in CNN, primarily due to the processing unit redesign and construction of new blocks[18]. After its introduction into and superior output over ImageNet in, Convolutional Neural Network (CNN) architectures have been widely used in image recognition tasks. More recently, given the additional difficulties of increasing image complexity and less labelled data the field of medical image analysis was successfully used. Some results include a segmentation of the brain tumor, diagnosis of diabetic retinopathy, identification of the nodule for pulmonary cancer, and an X-ray test for pneumonia and diagnosis of Alzheimer's [16].

Several economists support the usage of capital markets of neural networks and models of measurement of economic development. We concentrate on the forecast of the stock sector by analysing previous research [19]. A hybridised methodology has been proposed incorporating the usage of technological and fundamental criteria of bond market metrics in the estimation of daily stock prices. The research used multilayer perceptron models with back propagation algorithms (a basis for the neural feedforward network) [20]. The best outcomes of the two methods are contrasted (hybridised and scientific analysis). Scientific results found that the hybridised approach's degree of precision is higher than its technological form. In forecasting China's global index, and its results of study showed that the described model beats the regression model which used Stochastic Time Efficient Neural Networks. In predicting stock price index movement on the Istanbul Stock Exchange compared neural and ANN networks results. Technological metrics like CCI, MACD, RPM etc. are used in the input variables of the proposed models [21]. The findings demonstrated the accuracy of the neural path. His findings in the experiment revealed that networks perform better than the ANN methodology in prediction. A new model is being suggested to predict stock price in Shanghai. They used a neural network based on wavelet noise (WDBP) back propagation. The findings are related to the Back Propagation neural network and the overall tests revealed that the WDBP variant of the prediction index is greater than the BP model. The results indicate dominance of this experimental paradigm for prediction [22].

III. Methodology

A. Dataset

The dataset we are using is called Alzheimers Dataset (4 class of images), which is a part of the Kaggle dataset since the last seven months. The data was extracted in a .Jpg format. In this dataset, there were two folders which are train and test each containing four different classes of Dementia. These categories are Mild Demented, Moderate Demented, Non-Demented and Very Mild Demented which are the four classes to slot a patient.

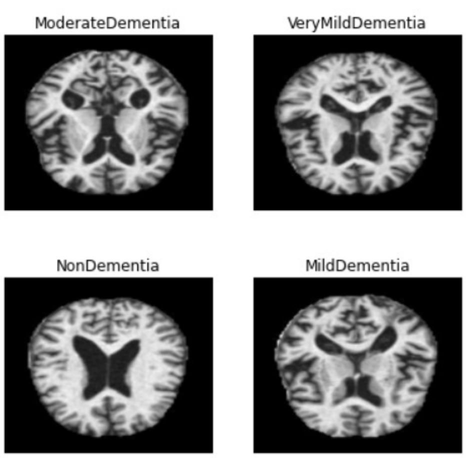


Fig. 1. Four classes of AD (MRI images)

Originally, the Data was hand collected from various websites with each and every label verified. The dataset consists of 6400 images of MRI segmentation which as we have explained earlier is one of the methods of detecting Alzheimers disease. Out of that the train folder consists of 5121 images which is segregated into four classes, i.e. 717 files are in Mild Demented class, 52 files in Moderate Demented, 2560 files in Non-Demented and 1792 files in Very Mild Demented. On the other hand the test folder consists of 1279 files which is nearly 20% of the dataset which is again segregated into four classes, i.e. 179 files are in Mild Demented class, 12 files in Moderate Demented, 640 files in Non-Demented and 448 files in Very Mild Demented.

The complete image classification task can be defined as follows:

Our input is a training dataset that consists of 5121 images, each labelled with one of 4 different classes. Then, we use this training set to train a classifier to learn what every one of the classes looks like. Finally, we test the classifier 's quality by challenging it to determine labels for a new collection of pictures that it had not previously encountered. We must then equate the actual labels of such pictures with the ones the trained model predicts

B. Data Pre-Processing Methods

1) Setting the path and Resizing the MRI image to 45x45: We used the glob.glob() function directly from the glob module to access paths recursively from within the directories and sub directories to remove the picture preprocessing dataset tab. We have used the OpenCV-Python library to load an image from the specified format, and resize the image using the OpenCV library cv2 function cv2.resize). As BGR is commonly used consistently throughout OpenCV so we had to convert all our images using the cv2.cvtColor() function.[6] This was done on both on train and test folders on all the four classes of Dementia.

2) Labeled the four different classes of MRI images: Then we gave label to our four classes which are 0: 'Mild Demented', 1: 'Moderate Demented', 2: 'Non Dementedd', 3:'Very Mild Demented' by using the id_to_label_dict method.

3) Splitting the data in the ratio of 70:30: Once we had pre-processed the images well, we splatted our dataset in the ratio of 70:30 which means 70% of the overall data was used for the training process and once the network was very well trained we tested it using the rest 30% of the data images. At the end we had our images dataset ready with the shape of (6400, 45, 45, 3).

4) Normalization: Normalization is a scaling method used in pre-processing stage to get a better prediction.After getting the training and testing set, we had to normalize the data images to the range of [0, 1]. Because the value of the RGB varies from 0 to 255, since it requires precisely one byte of data. One byte is equivalent to 8 bytes, and each bit indicates either a 0 or a 1 and we divide all train and test images by 255 to simplify the vector procedures.

5) Make a flattened version for some of our models: For making use of multi-layer neural network with five hidden layers and one having two dense layers we had to reshape both our training and testing data in order to feed the networks. Once that was done our flattened data shape was (4480, 6075),(1920, 6075)whereas our original data shape remained -(4480, 45, 45, 3) (1920, 45, 45, 3) (4480, 4) (1920, 4)for X_train, X_test, y_train., y_test respectively.

6) One Hot Encoding the Output:
Because we are working with categorical and noncontinuous data, we want to convert our model into one-hot encoding. One-hot encoding are a way for the model to understand that we're looking at categorical instead of continuous data. since that is the format required by Keras to perform multiclass classification. It transforms the integer to a sequence of all zeros at the integer index save for a 1. When the above step was finished, we had to hot encode our data by using the to categorical() function since we have multi-class results(03:4 output). In our case, this effectively transforms a class vector (integers) to a binary type matrix for categorical_crossentropy use[7].

C. Deep-learning Models

The second step is to fit the deep-learning classification models to a training set of the data and then evaluate them on a separate test set. In this project, we are using a few deep-learning models, such as Convolutional Neural Network, Multi-layer neural Network with five hidden layers, and Artificial Neural Network.

In a machine learning approach, based on the organization and working theory of biological neural systems,an idea emerged to Warren McCulloch and Walter Pitts when they analysed the mechanisms happening in the brain in 1943[8]. In addition, the Neural networks are composed of individual units called neurons. Nerve cells are nothing more than a series of classes-layers. Within each cell the neurons are connected towards the next layer's neurons. Information passes with certain compounds from the input layer to the output layer. A single node does a basic mathematical calculation. It then passes its data on to all the nodes that it is linked to.

1) Convolutional Neural Network:
Deep learning refers to neural nets with a large number of layers that derive a hierarchy of features from raw input images.Conventional machine learning techniques extract attributes manually, while in-depth learning extracts deep, high-level image features and trains vast volumes of data, leading to greater precision. Since considerably increased in the GPU processing power, deep learning methods allow us to train a vast amount of imaging data and increase accuracy despite variations in images.[9] Convolutional Neural Networks (CNN), introduced by Yann LeCun in 1988, is a unique model of artificial neural networks. CNN takes advantage of other functional brain functions. Object recognition is one of the most common uses of this architecture. Facebook , for example, uses CNN for automated tagging mechanisms, Amazon for product recommendations generation and Google for app search through among users photos.[10]

The key aspect in classifying images is recognition of the reference image and the corresponding class description.In our case the MRI image is passed via a series of convolutional, nonlinear, pooling layers and completely fully - connected and the output is then generated. The layer at Convolution is usually the lead. The image is fed into it, we should find it identical to the one which starts at the top left of the image when reading the input matrix. Then the algorithm chooses a narrower matrix there, also known as neuron, which is considered a buffer. The filter then induces convolution, i.e. passes around the representation of the inputs. The purpose of the filter is to multiply its values by the original pixel values. It sums up all such multiplications. In the end, one number is obtained. While the filter has already read the image in the upper left corner, 1 unit executing a similar operation goes forward and forward away.

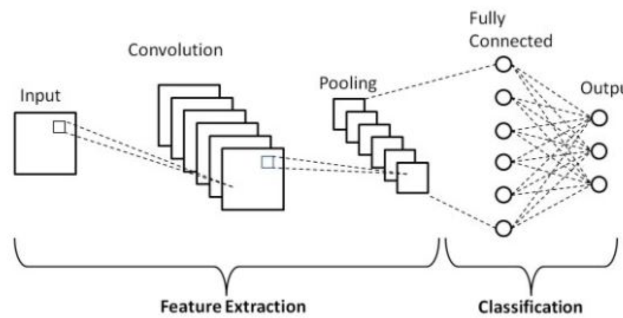


Fig. 2. Convolutional Neural Network (CNN) architecture

A matrix is formed after moving the filter over all positions, but smaller than a vector of inputs. From either a human viewpoint this process is similar to the recognition of boundaries and simple colours on the image. But in order to recognize the properties of a higher level such as the trunk or large ears the whole network is needed.[10]
The network will be composed of many convolutional networks combined with nonlinear layers and pooling. As the image moves through one convolution layer, the data for the second layer is the output of the first layer. And that happens for any additional convolutional layer.[10]After each process of the convolution the nonlinear layer is inserted. This layer has a feature to trigger which brings nonlinear properties. Without all this structure a network may not be large enough and cannot construct the response. In fact, the nonlinear layer is accompanied by the pooling layer. This deals for image width and height and does down-sampling on them. As a consequence, the width of the image is diminished. This ensures that if such attributes have been established in the typically done procedure, a complex image is no longer required for even more analysis and reduced to less complex pictures. If the sequence of convolutional, nonlinear, and pooling layers has been completed, adding a completely connected layer is of utmost importance. This layer may collect information from convolutional networks for the output. Assigning a fully

connected layer to an end of the network results in a N dimension matrix, where N is the number of classes the algorithm selects the required class from.

The cycle of reverse passing starts after the forward pass is complete. That's where CNN is collecting feedback and developing itself. That way the after-layer gets input from the corresponding layer, after the prediction is made. Such input during estimation would be in the form of losses suffered at each point. Now the CNN algorithm 's key aim is to achieve optimum losses. It is termed the local minima. Based on the input, the system is configured to change the kernel weights, which will increase the convolution efficiency as it occurs next time it moves forward. Likewise, failure will descend until the next forward pass occurs. Once we need to relay again, the network will start to change, there will be a further failure and repeated transmission. This forward pass followed by back propagation keeps happening the number of times we choose to train our model that is also termed as the epochs.

For our model, we have used the sequential model imported from Keras and Dense, Dropout, Flatten, Conv2D, MaxPooling2D as the layers imported from the Keras layers. Although the MRI databases used for Alhziemer 's diagnosis are usually very limited relative to the datasets used in image processing, training a deeper CNN model with a wide number of parameters was a big challenge for us. Below is the detailed description of the layers which we had used to train our network:

[1]Conv2D Layer This layer produces a convolution kernel, which is transformed to generate a tensor of outputs with the layer data. If use_bias is Valid, it generates and applies a bias function to the outputs. Eventually, if activation isn't Any, it would even refer to the outputs Since we are using this layer as the first layer in our model, we provided the keyword argument input_shape (tuple of integers, i.e. input_shape=(45, 45, 3) for 45x45 BGR MRI images in the dataset.[4]

[2]MaxPooling2D Layer This layer will be used to Down sample the input description by setting the maximum values for each dimension along the features axis over the window specified by pool_ size. Within each dimension the window is changed by measures. The output variable by using the padding option "real" has the following shape: output shape = (input shape-pool size + 1)/strides)[4]

[3]Dense Layer The Dense layer performs the operation: output = activation(dot(input, kernel) + bias) whereby activation is the element-wise activation function transmitted as the activation statement, kernel is a layer-generated weights matrix and bias is a layer-generated bias vector(only true if use bias is True).[4]

[4]Dropout Layer The Dropout layer dynamically sets input units to 0 at each stage during training period with a rate frequency which helps to prevent over-fitting.Inputs which are not set to 0 are multiplied by 1/(1 -rate), so that the sum is constant for all inputs. In addition to that, the Dropout layer only applies when training is set to True such that no values are dropped during inference. When we are using model.fit, training will be appropriately set to True automatically, and in other contexts, we are able to set the kwarg explicitly to True when calling the layer. [4]

2) Artificial Neural Network:

Artificial neural networks (ANNs) are architectures of mathematical learning motivated by the features of the biological neural networks. These are used for a wide variety of functions, ranging from fairly basic grouping challenges to voice recognition and Computer vision. ANNs are largely based on biological neural networks in the sense that they are implemented as an integrated computing device sometimes called nodes, which are functionally analogous to biological neurons.[11]

Therefore, at each layer they are a collection of several perceptrons. Ann's are also known as a Feed-Forward Neural network because inputs are transmitted in a forward direction only. some nonlinear mechanism can be learned by the Artificial Neural Network. Such networks are also popularly referred to as Fundamental Function Approximators. Activation functions incorporate nonlinear network properties. This lets the network understand certain dynamic input-output relationships[12].

The ties between various nodes have numerical values, called weights, and the network is gradually able to estimate the desired function by constantly altering those values. Every node in the network takes several inputs from many other nodes and determines one output based on the inputs and the weights of the connexions. In fact, this output is fed into another neuron, therefore repeating the whole process. Compatible with the details provided in the last phrase, one can quickly understand the artificial neural network's internal hierarchical structure, whereby neurons are divided into different layers. The input layer holds the signals, and an output layer. The particles lie between those are considered secret layers.[12]

It is important to think of the hidden layers as individual feature detectors, detecting increasingly complicated trends in the data as it is propagated across the network.For instance, If the network has an animal recognition task,then the first oversized layer will act as a line detector; the second oversided takes these lines as input to make a leg, the third oversided layer takes the leg and matches the leg with an eye, etc. This hierarchy helps the network to ultimately identify structures of great complexity[12].

As already stated, the network can precisely estimate an arbitrary function by systematically altering its weights. Initially, randomized values are given to the weights, and the network needs to be trained to find the weight parameters that have the desired result.In order to do this, we must first align the neural network output with the expected output, calculate the error and use this to change the network weight in line with the network's contribution to the output error. Neural networks are very slow learners and need a large amount of computing resources to achieve optimal results owing to the computationally costly back propagation process that needs to be performed for any series of inputs. However,one must realise that the top of the line networks currently comprises millions of neurons and up to a billion link weights in some serious situations.

Challenges with Artificial Neural Network (ANN)

While addressing a problem of image classification using ANN, the first step is to transform a 2-D image into a 1-D vector before the model can be equipped. It has the drawbacks listed below:

[1]The number of trainable parameters increases drastically with an increase in the size of the image. Therefore, due to this Unexplained behavior of the network reduces the overall trust in the network.

[2]ANN is missing the image's spatial characteristics. Spatial characteristics relate to pixel organisation in an image. The Vanishing and Exploding Gradient is one common problem in all those neural networks. This issue is related to the algorithm for back propagation. Via this back-propagation algorithm, a neural network's weights are modified by identifying the gradients.

[3]Thus, in the case of a very deep neural network with a very large number of hidden layers, the gradient disappears or bursts as it propagates downwards, allowing the gradient to vanish and burst, and ANN was unable to catch sequential information in the input data used to manage sequence data.

3) Multi-layer neural Network with five hidden layers: A multilayer perceptron (MLP) is the easiest form of feed forwards network. Many perceptrons with multilayers have really nothing to do with the original perceptron algorithm. The units are grouped in a series of layers here, and each layer comprises a certain set of equivalent units. Each unit with one layer is connected to each unit in the next layer; we state the network is linked in full.

The very first layer is the input layer, and the values of the input functions are taken from its units. The last layer is the output layer, and with each value the network outputs have one unit, which means a single unit for binary classification, and K units for K-classification. All the layers under these layers are classified as hidden layers, since we do not know in advance what these units should compute, and this needs to be discovered during learning. In addition to that, all activation functions are thresholds at zero and these layers are known as input units, output units, and hidden units. For our model we have used relu as the activation function.Also we discovered that the terminology for the depth is very inconsistent. A network with one hidden layer could be called a one-layer, two-layer, or three-layer network such as in our case we have a five-layer network since we have got five hidden layers.[13]To add to that for our neural network we have made use of the Dropout layers to remove features and additionally fight over-fitting

IV. Results and Analysis

After working extensively on the dataset, i.e. applying all the essential pre-processing steps as explained above we got the following results for our experiments:

1] For CNN when we trained the network on 4480 samples and validated on 1920 samples using the 2D Convolution Layer and Adadelta as our optimizer having a kernel_size of (3, 3) , relu and softmax as our activation functions, with 128 MRI images propagated through the network and with 20 passes over the entire dataset we got a Test accuracy of 0.90. in addition, we got a Test loss of 0.24 which started from 1.01 during the first epoch.

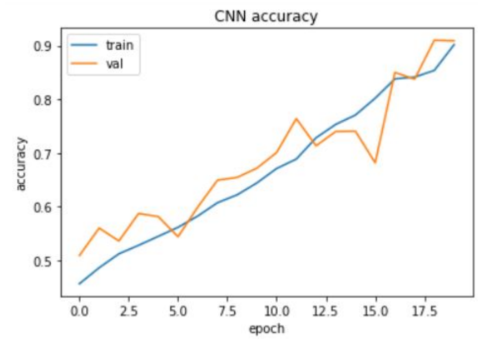


Fig. 3. CNN Accuracy Curve

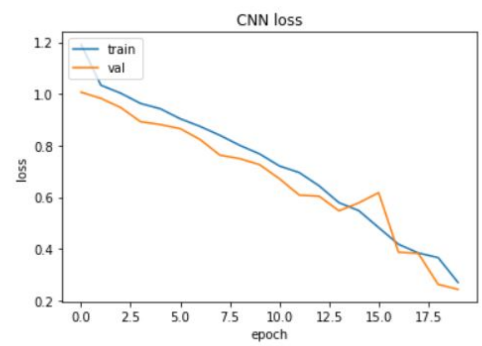


Fig. 4. CNN Loss Curve

Fig 3 and 4 displays the accuracy and loss curve for the CNN architecture. Our training accuracy increases with every epoch and reaches 0.88 and the validation accuracy is at 0.90.Fig 4 represents the curve of loss through our model run. Training run started with a higher loss of 1.19 to 0.26 through 20 epochs. While validation did quite well with 0.24 as its loss after 20 epochs.

2] For ANN when we trained the network on 3136 samples and validated on 1344 samples using relu and softmax as our activation functions and Adadelta as our optimizer, with 32 MRI images propagated through the network and with 50 passes over the entire dataset we got a Test accuracy of 0.77.

in addition, we got a Test loss of 0.63 which started from 0.99 during the first epoch.

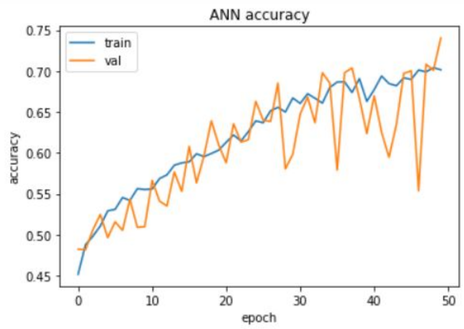


Fig. 5. ANN Accuracy Curve

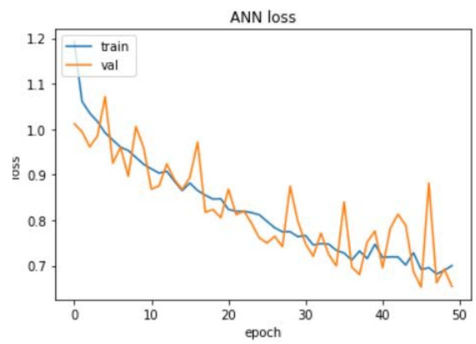
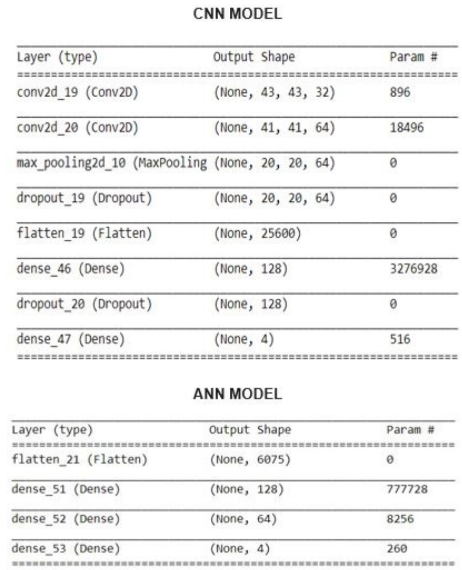


Fig. 6. ANN Loss Curve

Fig 5 and 6 displays the accuracy and loss curve for the ANN architecture. Our training accuracy increases with every epoch and reaches 0.75 and the validation accuracy is at 0.77.Fig 6 represents the curve of loss through our model run. Training run started with a higher loss of 1.03 to 0.56 through 50 epochs. While validation did moderately well with 0.63 as its loss after 50 epochs.Fig 7 shows the CNN and ANN model architecture which we obtained after tuning its parameters and thorough data evaluation.

3] For MLP we trained the network on 4480 samples and validated on 1920 samples using relu and softmax as our activation functions and RMSprop as our optimizer, with 32 MRI images propagated through the network and with 30 passes over the entire dataset we got a Test accuracy of 0.75. in addition, we got a Test loss of 0.60 which started from 0.99 during the first epoch.

Fig 8 displays the accuracy and loss curve for the MLP architecture. Our training accuracy increases with every epoch and reaches 0.72 and the validation accuracy is at 0.75.Moreover,Training run started with a higher loss of 1.10 to 0.67 through 30 epochs. While validation did comparatively well with 0.60 as its loss after 30 epochs.Fig 9 shows the MLP model architecture which we obtained after tuning its parameters and thorough data evaluation.



CNN MODEL

Layer (type)	Output Shape	Param #
conv2d_19 (Conv2D)	(None, 43, 43, 32)	896
conv2d_20 (Conv2D)	(None, 41, 41, 64)	18496
max_pooling2d_10 (MaxPooling	(None, 20, 20, 64)	0
dropout_19 (Dropout)	(None, 20, 20, 64)	0
flatten_19 (Flatten)	(None, 25600)	0
dense_46 (Dense)	(None, 128)	3276928
dropout_20 (Dropout)	(None, 128)	0
dense_47 (Dense)	(None, 4)	516

ANN MODEL

Layer (type)	Output Shape	Param #
flatten_21 (Flatten)	(None, 6075)	0
dense_51 (Dense)	(None, 128)	777728
dense_52 (Dense)	(None, 64)	8256
dense_53 (Dense)	(None, 4)	260

Fig. 7. CNN and ANN Model Architecture Summary

As shown in the table below,after our complete analysis on the dataset, we found out that CNN was the most optimum neural network for this classification task as it outperformed other neural networks by a vast margin,with a test accuracy of 0.90.To add to that,we also discovered that ANN and MLP performed quite similarly and gave us an accuracy score of around 0.77 and 0.75 respectively.However the Validation loss for both these neural networks is comparatively more when compared to the convolutional neural network.

Neural Network	Test_Accuracy	Val_Loss	Epochs
CNN	0.90	0.24	20
ANN	0.77	0.63	50
MLP	0.75	0.60	30

V. Conclusion

After performing the above-mentioned experiments we can conclude that our objective to find the most appropriate deep learning model for the AD classification was accomplished.it can be very well seen from the results that how CNN has outperformed other deep learning models with a test accuracy of 0.90.The results of these convolutional neural network will eventually help the doctors in providing appropriate Alzheimer's disease treatment to the patients based on the four different classes which we have considered in our experiment.Last but not the least we would just say that Presently our comprehension of the AD is stronger than ever and there are therapies that canhelp prolong the lives of sufferers and enhance their quality of life. With continuing clinical and translational research, we are be optimistic that drugs specifically addressing the disorder, and not only the effects, are on the horizon.

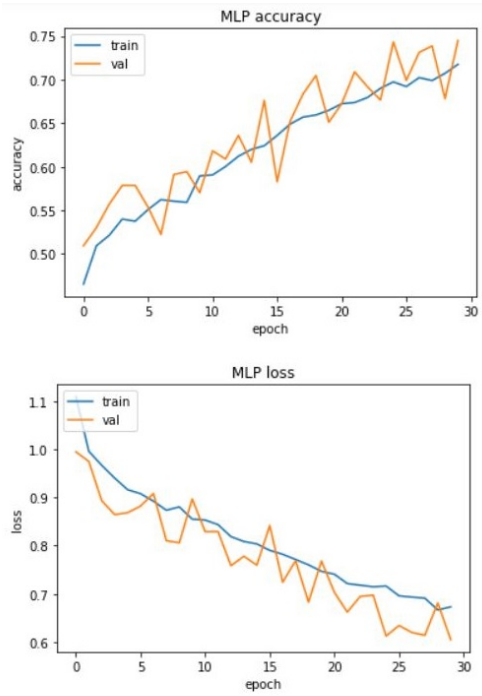


Fig. 8. MLP Accuracy and Loss curves

Layer (type)	Output Shape	Param #
dense_170 (Dense)	(None, 256)	1555456
dropout_126 (Dropout)	(None, 256)	0
dense_171 (Dense)	(None, 128)	32896
dropout_127 (Dropout)	(None, 128)	0
dense_172 (Dense)	(None, 128)	16512
dropout_128 (Dropout)	(None, 128)	0
dense_173 (Dense)	(None, 128)	16512
dropout_129 (Dropout)	(None, 128)	0
dense_174 (Dense)	(None, 128)	16512
dropout_130 (Dropout)	(None, 128)	0
dense_175 (Dense)	(None, 4)	516

Fig. 9. MLP Architecture Summary

VI. Research Contribution

For this research,our main concern was to find a suitable dataset on which we could work to clearly classify the different categories of the AD.After extensive search by both me and Rohit,We finally decided to go ahead with the Alzheimer's Classification dataset which I found on kaggle.Once that was done we needed to study the data in details and prepare it in such a way that different deep neural network models could be applied on it.Therefore we divided the pre-processing steps among ourselves.The first 3 steps of pre-processing as explained before in the report were done by Rohit and the rest were taken care by me. Now when our data was ready to explore, Rohit worked to fine-tune the ANN and in the meantime I focused all my energies in employing the convolutional neural networks(CNN). We also explored the RNN's but unfortunately we were unable to achieve good results,hence we decided mutually to drop that network.We then decided to go ahead with multi layer neural networks with 5 hidden layers as that gave us good results on the dataset.

As far as the project report is concerned Rohit worked on the introduction and literature review part whereas I focused on the methodology and results part of the report. In the end,I would just say that we have collaborated in a true sense to work on this research and it would not have been possible to complete this without each others constant motivation and support.

VII. Future Works

Researchers are recommended to try other neural networks including Inception Network, Residual Network and more recent state-of-the-art networks as a base model to construct the classifier for future research on this subject. Besides that, when selecting the model architecture and training hyper-parameters there are a wide range of choices. Although we have done our utmost to make meaningful choices and test a relatively large number of possibilities, we cannot exclude that other choices could have led to better results.

One might also seek to accomplish similar or better results by doing pre-processing steps such as stripping the skull, normalising the intensity. In addition, actual performance can also be improved through fine tuning, i.e. training the base model's pre-trained convolutional layers, provided the statistics used is in sufficient quantities and resources available can handle the increased computational burden. Moreover, using the neural impulses model, this approach can be further refined and all slices of an issue can be combined to obtain representative features. In addition, the same methodology can be evaluated for more samples and other sophisticated neural networks and machine-learning algorithms and optimisation methods on other volumetric and diagnostic functions.

References

[1] Miriam R., Malin B., Jan A., Patrik B., Fredrik B., Jorgen R., Ingrid L., Lennart B., Richard P., Martin R., et al. PET/MRI and PET/CT hybrid imaging of rectal cancerdescription and initial observations from the RECTOPET (Rectal Cancer trial on PET/MRI/CT) study. Cancer Imaging. 2019;19:52.

[2] Rebecca S., Diana M., Eric J., Choonsik L., Heather F., Michael F., Robert G., Randell K., Mark H., Douglas M., et al. Use of Diagnostic Imaging Studies and Associated Radiation Exposure for Patients Enrolled in Large Integrated Health Care Systems, 19962010. JAMA. 2012;307:24002409

[3] . Hsiao C.-J., Hing E., Ashman J. Trends in electronic health record system use among office-based physicians: United States, 20072012. Natl. Health Stat. Rep. 2014;1:118

[4] TensorFlow Core v2.3.0. 2020. Module: tf.keras.layers. https://www.tensorflow.org/api_docs/python/tf/keras/layers.

[5] Health, National Institutes of. n.d. Magnetic Resonance Imaging (MRI). https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri.

[6] Health, National Institutes of. n.d. Magnetic Resonance Imaging (MRI). https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri. Python Software Foundation. 2020. glob Unix style pathname pattern expansion. https://docs.python.org/3/library/glob.html.

[7] TensorFlow Core v2.3.0. 2020. Module: tf.keras.layers. https://www.tensorflow.org/api_docs/python/tf/keras/layers. . 2020. tf.keras.utils.to_categorical.

[8] Demuth, H.B., Beale, M.H., De Jess, O., Hagan, M.T.: Neural Network Design, Martin Hagan (2014)

[9] D. Rav et al., ”Deep Learning for Health Informatics,” in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 4-21, Jan. 2017, doi: 10.1109/JBHI.2016.2636665.

[10] Jain, R., Jain, N., Aggarwal, A. and Hemanth, D.J., 2019. Convolutional neural network based Alzheimers disease classification from magnetic resonance brain images. Cognitive Systems Research, 57, pp.147-159

[11] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016 Nov 10

[12] Yegnanarayana B. Artificial neural networks. PHI Learning Pvt. Ltd.; 2009 Jan 14.

[13] Grosse, Roger. 2019. ”Multilayer Perceptrons.” https://www.cs.toronto.edu/ mren/teach/csc411_19s/lec/lec10_notes1.pdf. Health, National Institutes of. n.d. Magnetic Resonance Imaging (MRI)

[14] D. Rav et al., ”Deep Learning for Health Informatics,” in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 1, pp. 4-21, Jan. 2017, doi: 10.1109/JBHI.2016.2636665.

[15] Jain, R., Jain, N., Aggarwal, A. and Hemanth, D.J., 2019. Convolutional neural network based Alzheimers disease classification from magnetic resonance brain images. Cognitive Systems Research, 57, pp.147-159.

[16] Justus D, Brennan J, Bonner S, McGough AS (2019) Predicting the Computational Cost of Deep Learning Models. In: Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018.

[17] Nguyen G, Dlugolinski S, Bobak M, et al (2019) Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. Artif Intell Rev 52:77124. doi: 10.1007/s10462-018-09679-z.

[18] Abbas Q, Ibrahim MEA, Jaffar MA (2019) A comprehensive review of recent advances on deep vision systems. Artif Intell Rev 52:3976. doi: 10.1007/s10462-018-9633-3.

[19] Qureshi AS, Khan A (2018) Adaptive Transfer Learning in Deep Neural Networks: Wind Power Prediction using Knowledge Transfer from Region to Region and Between Different Task Domains. arXiv Preprint arXiv181012611.

[20] Liu W, Wang Z, Liu X, et al (2017) A survey of deep neural network architectures and their applications. Neurocomputing 234:1126. doi: 10.1016/j.neucom.2016.12.038.

[21] Zhang Q, Zhang M, Chen T, et al (2019) Recent advances in convolutional neural network acceleration. Neurocomputing 323:3751. doi: 10.1016/j.neucom.2018.09.038.

[22] Ahmed U, Khan A, Khan SH, et al (2019) Transfer Learning and Meta Classification Based Deep Churn Prediction System for Telecom Industry. 110.