# Alzheimer's Disease Prediction Using Convolutional Neural Network Models Leveraging Pre-existing Architecture and Transfer Learning

Mahjabeen Tamanna Abed[1], Umme Fatema[2], Shanewas Ahmed Nabil[3],
Md. Ashraful Alam[4], and Md Tanzim Reza[5]
[1,2,3,4,5]Department of Computer Science and Engineering, BRAC University, 66 Mohakhali, Dhaka 1212, Bangladesh
Email: [1]mahjabeen.tamanna.abed@g.bracu.ac.bd, [2]umme.fatema@g.bracu.ac.bd, [3]shanewasahmed@gmail.com,
[4]ashraful.alam@bracu.ac.bd, [5]rezatanzim@gmail.com

***Abstract*—Early Alzheimer's Disease (AD) or Mild Cognitive Impairment (MCI) can be diagnosed through proper examination of several brain biomarkers. In recent times, several high-dimensional classification techniques have been suggested to discriminate between AD and MCI on the basis of T1-weighted MRI of patients. These techniques have been implemented mostly from scratch, making it really difficult to achieve any meaningful result within a short span of time. Therefore, the classification of AD is usually a very daunting and time-consuming task. In our study, we trained high dimensional Deep Neural Network (DNN) models with transfer learning in order to achieve meaningful results very quickly in terms of detecting AD from fMRI image. The fMRI image dataset has been collected from Alzheimer's Disease Neuroimaging Initiative (ADNI). We have used three different DNN models for our study: VGG19, Inception v3, and ResNet50 to classify AD, MCI, and Cognitively Normal (CN) patients. Firstly, we implemented some pre-processing steps on the images and divided them into training, testing, and validation sets. Secondly, we initialized these DNN models with the weights from pre-existing models trained on the ImageNet dataset. Finally, we trained and evaluated all the DNN models. After a relatively short amount of training (15 epochs), we achieved an approximate of 90% accuracy with VGG19, 85% accuracy with Inception v3, and 70% with ResNet50. Thus, we achieved excellent classification accuracy in a very short time with our research.**

***Contribution* — Classification between early-stage and late-stage AD at improved accuracy with transfer learning.**

***Index Terms*—VGG19, Inception, Residual Network (ResNet), Convolutional Neural Network (CNN), Transfer Learning, Magnetic Resonance Imaging (MRI)**

## I. Introduction

Alzheimer's disease is an irreparable disease of the brain which influences an individual's memory, thinking, and different capacities. It is basically a chronic neurodegenerative disease which is the cause of 60-70% of dementia issues. It is a brain-related disorder that culminates in loss of memory, cognitive, and intellectual disagreement that can adversely affect social activity and decision-making.

In the United States, it is the sixth most important reason for death, accounting for 3.6% of all fatalities in 2014 [1]. Currently, in Bangladesh, there is no accurate epidemiological data on AD. However, some sparse information on the number of AD patients in Bangladesh is accessible [2]. As a developing country, it is not yet equipped to handle AD as the knowledge of AD is not sufficient yet. All in all, there are many people throughout the world who are suffering from this disease and often the disease becomes fatal. As a result, important preventive measures are needed to be taken.

### A. Motivation

The Alzheimer's Disease Research (ADR) program has been granted almost 120 million US dollars since 1985 to promote promising research in areas of molecular biology, genetics, and epidemiology. In 2013, 84767 deaths from AD were reported in statutory death certificates. The actual amount of casualties related to AD (or deaths from AD) is estimated to be much higher than the number of fatalities from AD published on official documents in the U.S [3]. In Bangladesh, the existence of Alzheimer's disease is not as alarming as other nations especially the United States. However, After 20 to 30 years our young generation would become older which consists of a major part of our total population [4]. Additionally, the number of smokers and people with obesity which contributes to causing Alzheimer is increasing. Therefore, in Bangladesh, there is a huge probability that Alzheimer's will be a threatening disease in the near future. The motivation behind choosing this topic as our research topic is to try to predetermine this disease as it is threatening and alarming throughout the globe. Additionally, the motivation is to distinguish the features between Alzheimer's disease and MCI as it this difficult to differentiate the features of these two diseases.

### B. Contribution

Firstly, our contribution through this research is distinguishing between early-stage and late stage. We have distinguished the symptoms between AD and MCI and differentiated AD from MCI through early biomarker detection during the whole process. Our accuracy rate is improved as we have used transfer learning. Transfer learning prevents our model from getting stuck into local optima, ensuring decent results with minimum effort.

## II. Background Study

### A. Previous Works

In recent years, significant progress in neuro-imaging has given the possibilities for the control of neurological diseases, resulting in improved early and accurate identification of AD [5], [6]. In genetic research, machine learning models were used to explore genetic variants that are most associated with different complex diseases [7]. Due to its non-invasive nature and absence of pain for patients, Magnetic Resonance Imaging (MRI) is also used in AD-related research. MRI has attracted significant interest as a tool for identifying biomarkers of Alzheimer's disease [7]. Additionally, Magnetic Resonance Imaging provides a spatial resolution and smart distinction [8]–[10]. These developments have enormous potential for slow realization of the technology in medical imaging, analysis of medical data, medical diagnostics, and general health care. Another significant region of implementation is the sophisticated registration of deformable images, allowing quantitative analysis across distinct modalities of physical imaging over time [10]. Many kinds of research have therefore used structural MRI (sMRI) based biomarkers to classify AD [11], [12], explaining brain atrophy and altering the volume of brain tissues. Similarly, functional Magnetic Resonance Imaging (fMRI) [13] is often used to characterize the hemodynamic response related to neural activity and functional or structural properties, which could be used to define medical anomalies such as abnormalities in the entire brain at the point of connectivity [14], [15].

### B. CNN architecture

In this research, we have used different models of Convolutional Neural Network (CNN) for the classification of AD. A Convolutional Neural Network, also recognized as CNN or ConvNet, is a category of neural networks that are specialized in grid-like topology cognition, like those of pictures. An image that is in digital form is a discrete portrayal of visual information. CNN has introduced a potent system to solve the challenges with image recognition [16]. These are influential methods that can operate dynamically on the given inputs [17]. It includes a grid-like pattern of pixels that includes numerical values to indicate how luminous every pixel must be and what pigment it must be. The basic feed-forward neural networks can act as a powerful classification tool that can classify complex images. However, all the nodes here are connected to each other, and having connections from all nodes in one layer to all nodes in the next layer is extremely inefficient. A cautious reduction of the links based on domain expertise, i.e. the image composition, results in much better performance [18]. CNNs allow data-driven, extremely representative, hierarchical image functions to be learned from required training data. Currently, there are three main methods of using CNNs adequately to distinguish medical images: train the CNN from scrap, use CNN attributes off-shelf, and monitor unsupervised pre-training with supervised adjustment. There are other efficient techniques such as transferring previously learned knowledge to medical image functions, i.e. pre-trained CNN models from the existing image data set with proper adjustments [19]. All these CNN models extract features dynamically from image input and then classifies it with the classification layers attached to the top of the feature extractor.

## III. Algorithms

For our proposed research, we used three models- VGG19, Inception V3, ResNet50. VGG-19 is a convolutional neural network that managed to properly classify more than one million pictures from ImageNet. The '19' in VGG stands for the number of convolution layers in the network. VGG network is the idea of a much deeper network with much smaller filters. Is has another variant named VGG16 which has 16 convolution layers as the name implies. As the filter is very small, only three by three convolution layers with the periodic pooling all the way which is basically the smallest filter size. The stack of three by three convolutional layers with stride 1 has the same response field as one seven by seven convolutional layers but deeper, has more non-linearities, and fewer parameters as well.

GoogleNet developed a module called the Inception module, which computes a narrow CNN with a compact standard shape. In a full GoogleNet architecture, there are Stem Network: convolutional pool-2 convolutional pool, stack inception layers with reduced measurements then at the edge of classifier output. The fully connected layers have been removed. It has parameters that are twelve times lower than AlexNet. It is a broader computer-efficient network. It substituted the fully connected layers with a basic average global pooling at the concluding part (Auxiliary classifications output to inject additional gradient at lower layers) in which the channel values are being averaged over the 2D feature map whenever the end of the convolution layer is reached. The auxiliary layers are actually useful for the final classification because when they are being trained they do average all these for the losses coming out. All of the layers have separate weights.

ResNet has very deep networks, much deeper than any other network using residual connections. They had a 152 layers model for ImageNet.They were able to get 3.5 of 7 percent top of the error. The deeper designs usually perform worse, but Resnet is different. ResNet creators' hypothesis is that a deeper model is more difficult to optimize. However, deeper models are needed for complex classification so skip connections were added to avoid overfitting. It is a classic neural network used in many computer vision functions as a backbone. ResNet is a type of unique architecture based on modules of micro-architecture which is also called "network-in-network architectures".

## IV. Dataset Details and Processing

### A. Dataset Description

In our research, we collected a dataset from Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI scientists accumulate, formalize, and use information as predictors of the disease, including pictures of MRI and PET, genetics,

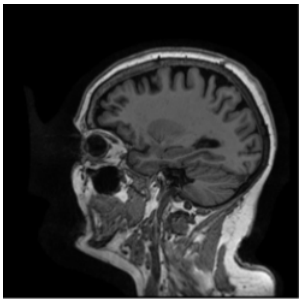

Fig. 1. Single slice of a cognitively normal MRI image

cognitive tests, CSF, and blood biomarkers. They provide research tools and information from the North American ADNI research, inclusive of patients with AD, mild topics of cognitive impairment, and aged controls. During their involvement in the research, ADNI scientists obtain several kinds of information from research assistants, using a normal set of protocols and processes to eradicate incompatibilities. The LONI Image and Data Archive (IDA) allocates this data to approved researchers free of charge. We used this dataset for our research as this is an open-source dataset. We have collected fMRI of 54 patients for our research.

TABLE I
METADATA OF THE DATASET

| Accuracy | | | |
|---|---|---|---|
| Group | Sex | Age | Visit |
| AD | M | 80 | 2 |
| AD | M | 82 | 6 |
| AD | M | 80 | 2 |
| AD | M | 81 | 4 |
| AD | M | 57 | 2 |
| AD | M | 58 | 4 |
| AD | M | 57 | 2 |
| AD | M | 59 | 6 |
| CN | F | 78 | 2 |
| CN | F | 79 | 4 |
| CN | F | 78 | 2 |
| CN | F | 80 | 6 |
| MCI | M | 80 | 2 |
| MCI | M | 82 | 6 |
| MCI | M | 81 | 4 |
| MCI | M | 80 | 2 |
| CN | F | 79 | 4 |

### *B. Dataset processing methods*

1. Categorize NIfTI file: NIfTI (Neuroimaging Informatics Technology Initiative) is a working group sponsored by the NIH to support the scalability of functional software instruments for neuroimaging. NIfTI utilizes one or two formats for storing files. In a couple of documents, the dual format keeps information: a header file (.hdr) with metadata and a data file (.img). The single file format keeps the data in a single file (.nii), which includes data-followed header information. We organized and tagged the file according to patient-specific ID. We wrote a python code to execute this process.

2. Segmentation of 3D files: We used Python package med2image version 1.1.2. Med2image is a convenient Python 3 utility that converts formatted medical image documents to more visually adaptable files like png and jpg. The input content of 3D and 4D NIfTI is acknowledged. In the case of 4D NIfTI, in conjunction with a specific slice index, a specific frame can be clarified. In most instances, as most of the NIfTI data is 3D, only one slice is mandated. In addition, all slices, or just the middle one, can be converted. We converted NIfTI 3D to 2D jpg files and organized those according to user-specific ID.

3. Extracting and attaching the image extension: We wrote a Python code to extract images from specific IDs and separate the image object from the extension. Then we modified the object using the incremental ID and reattaching the extension. Therefore we could clean the image dataset. And delete the unnecessary segmentation of images.

4. Grouping into classes: We divided the data into three classes: AD, MCI, and CN. This process is done to distinguish features between AD and MCI. Because there are huge similarities between the features of AD and MCI MRI images. The Cognitively Normal class is created for comparing the normal subjects with the disease affected subjects. We divided the subjects into those groups against their user ID.

5. Train-Test split: The train and test part is divided into 80 percent and 20 percent accordingly. We distributed this in a uniform process. Because if we just split them into random 80 percent and 20 percent then the distribution would not be uniformed. To overcome this, we took five subjects at each steps, four is taken to the training set and the rest one is given to the test set. Then we kept on repeating this step. That is how the subjects were split into train and test sets.

## V. PROPOSED METHOD

### *A. Structural MRI data acquisition*

The proposed model has been applied for the classification of both binary and multi-class image labels. We had a folder with all the pictures we wanted to train your model on. Then, we also needed real image labels to train the model. We also had a .csv file containing the names of all the training images and their respective true labels. We acquired a dataset from the Alzheimer's Disease Neuroimaging Initiative program which is an organization to share Alzheimer's research data with the world.

### *B. Data pre-processing*

The pre-processing module involves the standardization of inter-slice intensity, reduction of noise, and correction of non-uniformity intensity. Firstly, we loaded all the images and pre-processed them according to the requirement of our project. We generated a validation set to verify how our model performs on unseen information (test data). We trained and validated our model on the training set using the validation. Then we had saved images in an array file and checked if the image is valid or not. For invalid images, we had created a garbage collector. Using this, we removed the corrupted or

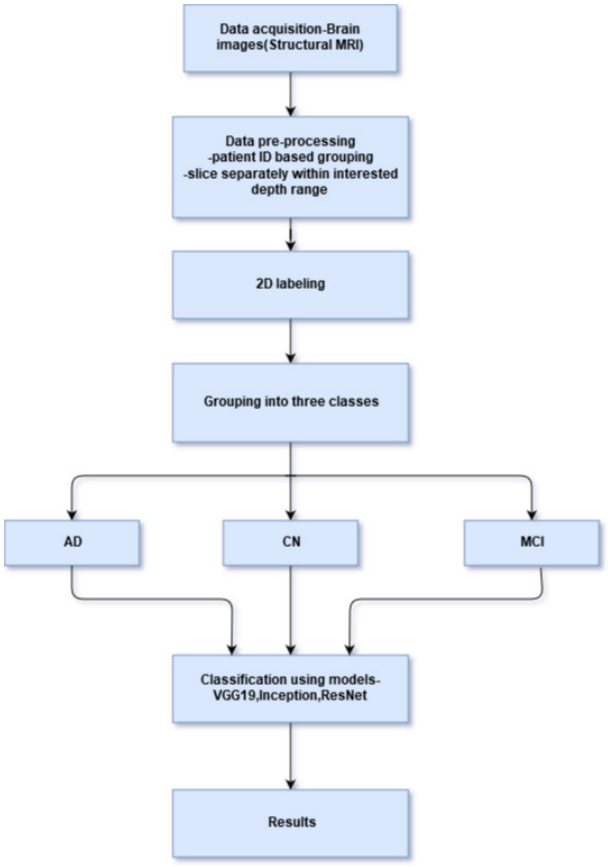


Fig. 2. Proposed Model

the unreadable images and the images that are not useful for example, parts of image slices of the neck portion that are not included in brain images. We filtered the data that has the most recent acquisition date along with a filter consisting of 3 Tesla magnetic resonance and a slice thickness of 2mm. The MRI image we collected was a structural magnetic resonance image and the format was in NIfTI. NIfTI is an analyzer-style data format which is the volume of 2D slices. We organized and grouped the data by specific patient identification. After doing so, we had three different classes of data which were cognitively normal, AD, and Mild Cognitive Impairment (MCI). Hence one class is inefficient to convert an MRI NIfTI image 2D volume to 2D image slice.

### C. Grouping into classes

In the next part, we separated the data of all the three classes of our dataset. Afterward, we extracted all the MRI slices individually so that these individual slices can be processed as a 2D image. The technique involved med2image, a python package that can slice medical scans into 2d image arrays. We got 180 image slices from one NIfTI file. From those 180 images per patient ID we dropped 20 slices from the top and 20 slices from the bottom as those slices do not really carry any meaningful information. That remaining 140 images are our main interest. In the next part of the process, we prepared the entire dataset to make it suitable for analyzing with machine learning.

In this process, we sequentially tagged each image object and separated those from image extension. After completion, we reattached the extension. Completing all the processes stated above has given us three unanimous groups of sequentially tagged and organized data. To clean and enhance those processed data we used image processing. We read each and every image to identify its validity and incorporation with our research domain. If we had anything which was unreadable or did not represent the region we featured of a human brain, we dumped that image and moved to the next one. After this crude process, we had a 0.3% loss of data. We labeled the 3 classes according to the disease category and control set. That gave us a full set of dataset classes ready and processed to start run through the CNN model using transfer learning. Therefore we split those 3 classes each into 2 separate sets as a training set and testing set. The split was 80% and 20%. A simple python code snippet used to perform this split.

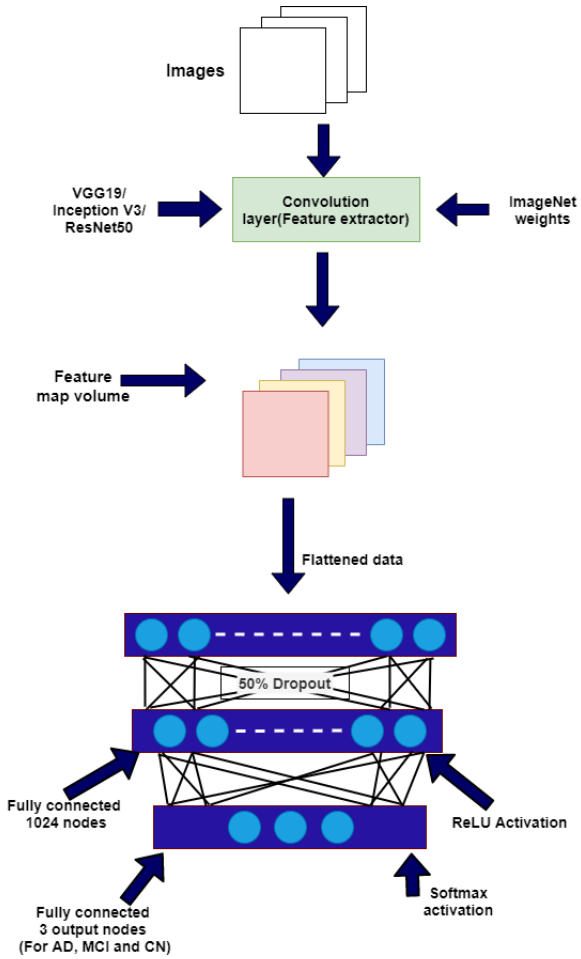


Fig. 3. Classification using CNN models

### D. Classification using CNN models

In our research, through transfer learning, namely VGG16, Inception, and ResNet, we explore the adaptation of three

common CNN architectures to an AD diagnosis issue. The architectures are easily accessible along with the pre-trained weights associated with it. While the architectures are being trained in a distinct domain, we demonstrated that pre-trained weights are very well generalized for AD diagnosis when the training data is intelligently selected. Here, VGG19 is 19 layered networks. By pushing up to 16-19 layers and using very tiny (3x3) convolution filters, it can investigate network depth. The architecture of Inception is a variant of Google's built deep learning architecture. There have been distinct iterations over the years. The break-through of the Inceptions is in the realization that it is possible to learn non-linear functions by altering how convolutional layers are linked. The fully connected layer is therefore removed in preference to a worldwide average pooling that averages the function maps and then connects for classification with a softmax layer. There are therefore fewer parameters, resulting in less overfitting. Residual neural networks have shown the chance of significantly enhancing the network depth while having rapid convergence.

After the pre-processing procedure, the dataset was used to train different types of CNNs: VGG19, Inception V3, and ResNet50. We tried to keep the whole structure as close as possible to the original structure. Therefore, before passing the input images through the CNNs, the images were compressed into (224,224) size which is the size that was used for the original CNNs trained on the ImageNet dataset. Additionally, we also used the weights that were calculated for the ImageNet dataset. The target was to achieve convergence as fast as possible. We used Keras which is the high-level API of TensorFlow to create and train models of deep learning.

## VI. Result and Analysis

We used ReLU activation functions throughout multiple models. We ran AD vs CN through VGG19 and got 78% validation accuracy on the first five iterations, on the next five got 89% and 3rd validation accuracy was 98%. On the other hand Inception v3 gave 83% on the first five iterations. The next two sets of iteration gave 86% and 95%.

Afterward, we ran VGG19 and Inception V3 through our second set, which contains AD vs MCI vs CN. Firstly VGG19 had 84% on the first five iterations, after that next five iterations gave 89% and last five had 93%, which is the highest validation accuracy we achieved through our three-class model. Moreover, Inception V3 had 83% validation accuracy, 85% for the next five iterations, and 89% accuracy achieved for the last set. Similarly, in ResNet, for AD Vs CN, the validation accuracy was 74%, 81%, and 85%. On the other hand for AD vs. MCI Vs CN result was 59%, 67%, 70 %.

From our results, we can state that we achieved an excellent result in our models. Our decision to drop out the multi-section layer from the image set had a great impact on the transfer learning process. Though we had lower accuracy with the Inception V3 model, our train and test graph shows the learning curve. In the graph, the loss was decreasing throughout the model while model accuracy was getting higher. We

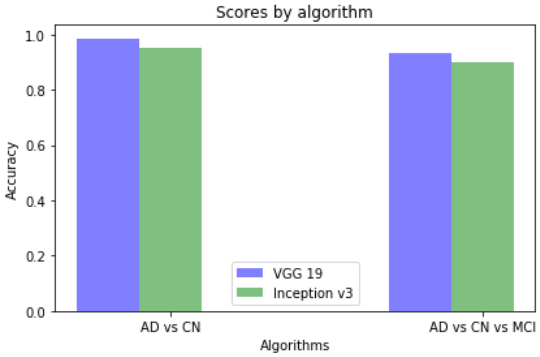


Fig. 4. Comparison between results VGG19 and Inception v3

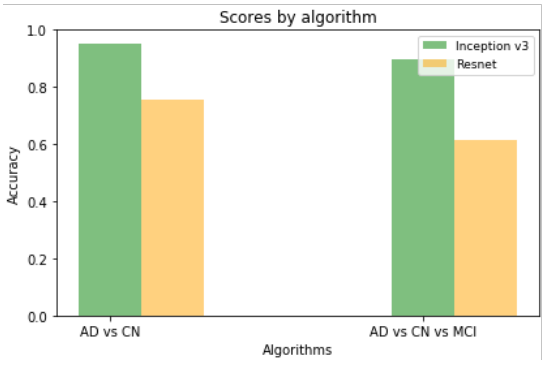


Fig. 5. Comparison between results from ResNEt50 and Inception v3

ran 15 epochs and stopped early cause the dataset was limited to 50 thousand image slices and we could eliminate the chances of overfitting. If we had more data incorporated with our research domain, our precision level could increase significantly. Our validation accuracy was higher than categorical accuracy which contrasts general machine learning concepts. As generally, training accuracy should be higher than test accuracy. However, the results differ in our simulation because we used a 50% dropout in the only fully connected layer. Therefore, during the training task, the model could not fully be used in its potential, since half of the weight was missing

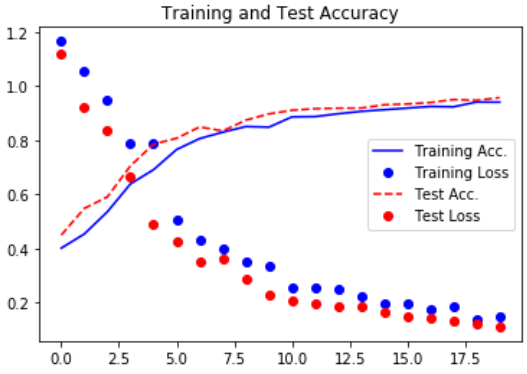


Fig. 6. Train Vs Test Accuracy Result

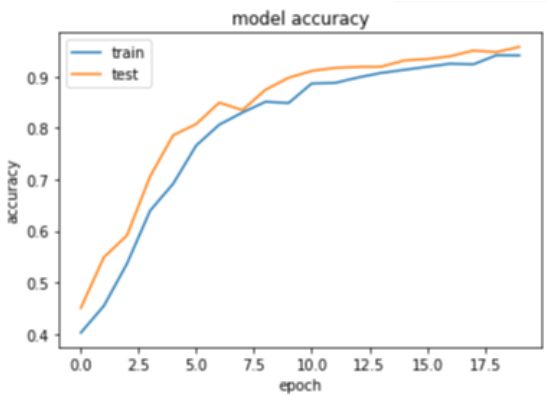


Fig. 7. Model Accuracy

during each iteration. From fig 7 we can see that the model improve significantly during the first 10 epoch or so. However, the improvement became less and less noticeable after the 10th epoch. So the training was stopped after the 15th epoch, as the improvement per epoch could be ignored. However, due to lack of time, a single split was made and for that, the results may be subject to a large statistical error.

## VII. Conclusion and Future Works

To conclude, from our experiment we can see that deep learning can work very efficiently for the classification of Alzheimer's disease from MRI images. Deep neural networks are certainly very effective when it comes to generating precise choices based on complex datasets. So, this is very understandable, that deep learning has a very crude way to solve a problem and give dynamic results towards the research domain. The involvement of deep learning can be extremely beneficial since it is not prone to human-made mistakes and can automate the task for the neurologists.

In our research, we have collected a structural MRI image dataset from ADNI and then sliced and processed the image data to make it usable for our model. We have studied the brain biomarkers of Alzheimer's disease for extracting the main features of the MRI images. However, when we have studied deeply, we have come to know that Alzheimer's also has an impact on the eye. The Alzheimer affected people's eye region can be differentiated from cognitively normal people's eye regions. The genetic biomarkers that are responsible for AD were also selected and categorized. For this process, we will be needing an OCT image dataset of AD affected people which is currently not available. In the near future, we hope that the dataset will be generated and we will continue our research on this area as eye biomarkers are an important perspective in Alzheimer's research. Hopefully, early detection will be more feasible and the accuracy rate will be much than this. This is our future work plan if we can manage the OCT fundus image dataset which is not available currently. Also, we will work with more recent models such as Inception V4, InceptionResNet, and DenseNet.

## References


[1] Jiaquan Xu, Kenneth D Kochanek, Sherry L Murphy, and Betzaida Tejada-Vera. Deaths: final data for 2014. 2016.

[2] Aminur Rahman, Farhana Salam, Md Aminul Islam, AKM Anwarullah, Md Rafiqul Islam, Md Nurul Amin Miah, Uttam Kumar Saha, and Zahed Ali. Alzheimer s disease-an update. *Bangladesh Journal of Neuroscience*, 28(1):52–58, 2012.

[3] Association Alzheimer's. 2015 alzheimer's disease facts and figures. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 11(3):332, 2015.

[4] Md Rahman, Afsana Tajmim, Mohammad Ali, Mostakim Sharif, et al. Overview and current status of alzheimer's disease in bangladesh. *Journal of Alzheimer's disease reports*, 1(1):27–42, 2017.

[5] Haruo Hanyu, Tomohiko Sato, Kentaro Hirao, Hidekazu Kanetaka, Toshihiko Iwamoto, and Kiyoshi Koizumi. The progression of cognitive deterioration and regional cerebral blood flow patterns in alzheimer's disease: a longitudinal spect study. *Journal of the neurological sciences*, 290(1-2):96–101, 2010.

[6] Katherine R Gray, Robin Wolz, Rolf A Heckemann, Paul Aljabar, Alexander Hammers, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative, et al. Multi-region analysis of longitudinal fdg-pet for the classification of alzheimer's disease. *NeuroImage*, 60(1):221–229, 2012.

[7] N Schuff, N Woerner, L Boreta, T Kornfield, LM Shaw, JQ Trojanowski, PM Thompson, CR Jack Jr, MW Weiner, and Alzheimer's; Disease Neuroimaging Initiative. Mri of hippocampal volume loss in early alzheimer's disease in relation to apoe genotype and biomarkers. *Brain*, 132(4):1067–1077, 2009.

[8] JM Górriz, Fermín Segovia, Javier Ramírez, A Lassl, and Diego Salas-Gonzalez. Gmm based spect image classification for the diagnosis of alzheimer's disease. *Applied Soft Computing*, 11(2):2313–2325, 2011.

[9] Kenji Suzuki. Overview of deep learning in medical imaging. *Radiological physics and technology*, 10(3):257–273, 2017.

[10] Iman Beheshti, Hasan Demirel, Alzheimer's Disease Neuroimaging Initiative, et al. Probability distribution function-based classification of structural mri for the detection of alzheimer's disease. *Computers in biology and medicine*, 64:208–216, 2015.

[11] George A Papakostas, Alexandre Savio, Manuel Graña, and Vassilis G Kaburlasos. A lattice computing approach to alzheimer's disease computer assisted diagnosis based on mri data. *Neurocomputing*, 150:37–42, 2015.

[12] Yudong Zhang, Zhengchao Dong, Preetha Phillips, Shuihua Wang, Genlin Ji, Jiquan Yang, and Ti-Fei Yuan. Detection of subjects and brain regions related to alzheimer's disease using 3d mri scans based on eigenbrain and machine learning. *Frontiers in computational neuroscience*, 9:66, 2015.

[13] Anders H Andersen, William S Rayens, Yushu Liu, and Charles D Smith. Partial least squares for discrimination in fmri data. *Magnetic resonance imaging*, 30(3):446–452, 2012.

[14] Yong Fan, Susan M Resnick, Xiaoying Wu, and Christos Davatzikos. Structural and functional biomarkers of prodromal alzheimer's disease: a high-dimensional pattern classification study. *Neuroimage*, 41(2):277–285, 2008.

[15] M Graña, M Termenon, A Savio, A Gonzalez-Pinto, J Echeveste, JM Pérez, and A Besga. Computer aided diagnosis system for alzheimer disease using brain diffusion tensor imaging features selected by pearson's correlation. *Neuroscience letters*, 502(3):225–229, 2011.

[16] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[17] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):221–231, 2012.

[18] Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2):102–127, 2019.

[19] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016.