Early Prediction of Alzheimer's Disease Using Convolutional Neural Network

by

Mahjabeen Tamanna Abed
17101268
Shanewas Ahmed Nabil
12201087
Umme Fatema
16101330

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
August 2019

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Mahjabeen Tamanna Abed
17101268

Shanewas Ahmed Nabil
12201087

Umme Fatema
16101330

Approval

The thesis/project titled "Early Predicting of Alzheimer's Disease Uisng Convolutional Neural Network" submitted by

1. Mahjabeen Tamanna Abed (17101268)
2. Shanewas Ahmed Nabil (12201087)
3. Umme Fatema (16101330)

Of Summer, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on August 07, 2019.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Ashraful Alam
Assistant Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Jia Uddin
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Mahbub Alam Majumdar
Professor
Department of Computer Science and Engineering
Brac University

Ethics Statement (Optional)

This is optional, if you don't have an ethics statement then omit this page

Abstract

Neuroimaging can be a prospective instrument for the diagnosis of Mild Cognitive Impairment (MCI) along with its more severe stage, Alzheimer's disease (AD). High-dimensional classification methods have been commonly used to explore Magnetic Resonance Imaging (MRI) for automatic classification of neurodegenerative diseases like AD and MCI. Early AD or MCI can be diagnosed through proper examination of several brain biomarkers such as Cerebrospinal Fluid (CSF), Media Temporal Lobe atrophy (MTL) and so on. Abnormal concentrations of the mentioned biomarkers on MRI images can be a potential sign of AD or MCI. In the 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, 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. We have used three different DNN models for our study: VGG19, Inception v3 and ResNet50 to classify between 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 relatively short amount of trainings (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.

Keywords:
Alzheimer's Disease(AD), VGG19, Inception, Residual Network(ResNet), Convolutinal Neural Network(CNN), Transfer Learning, Mild Cognitive Impairment(MCI), Magnetic Resonance Imaging(MRI)

Dedication

We would like to dedicate this thesis to our loving parents.

Acknowledgement

First and foremost, we are grateful to the God for the good health and wellbeing that were necessary to complete this book.
We wish to express our sincere thanks to Mahbub Alam Majumdar, Professor and Chairperson of the Computer Science and Engineering department, for providing us with all the necessary facilities for the research.
Foremost, we would like to express our sincere gratitude to our respectable supervisor Dr. Md. Ashraful Alam, Assistant Professor of BRAC University, Computer Science and Engineering department for the continuous support of our study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped us in all the time of research and writing of this thesis. We could not have imagined having a better advisor and mentor for our thesis. Without his assistance and dedicated involvement in every step throughout the process, this paper would have never been accomplished.
We are also grateful to all the lecturers in the Department. We are extremely thankful and indebted to them for sharing expertise, and sincere and valuable guidance and encouragement extended to us.
We would also like to thanks our Research Assistant Tanzim Reza for his supporting hand and proper guideline towards us.
I take this opportunity to express gratitude towards the Computer Vision and Intelligence System (CVIS) lab and Lab Technical officer (LTO) for providing the necessary facilities for the research. I also thank my parents for the unceasing encouragement, support and attention. I also place on record, my sense of gratitude to one and all, who directly or indirectly, have extended their hand in this venture.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AD Alzheimer Disease

$ADNI$ Alzheimer's Disease Neuroimaging Initiative

ADR Alzheimer's Disease Research

$APOE$ Apolipoprotein E

BP Blood Pressure

CN Cognitively Normal

CNN Convolutional Neural Network

CT Computer Tomography

DNN Deep Neural Network

FC Filter Concentration

FC Fully Connected

$FMRI$ Functional Magnetic resonance imaging

HF Hippocampal Formation

MCI Mild Cognitive Impairment

$NIfTI$ Neuroimaging Informatics Technology Initiative

PD Peritoneal Dialysis

$ReLu$ Rectified Linear Unit

$ResNet$ Residual Networks

$SMRI$ Structural Magnetic resonance imaging

SNR Substantia Nigra pars Reticulate

VGG Visual Geometry Group

Chapter 1

Introduction

Alzheimer's disease is an irreparable malady of the brain which influences an individual's memory, thinking, and different capacities. It is basically a chronic neuro degenerative disease which is the cause of 60-70% cause of dementia. It is a brain disorder that originates loss of memory, cognitive and intellectual disfigurement that can adversely effect social activity and decision-making. The occurrence of AD in females has been observed to be greater than in males. It is not apparent whether this difference is due to biology, the assumption that feminine appears to stay longer or their personalities. On the contrary, research from proofs towards a sex distinction within the danger of AD. Vascular dementia is additional frequent in males compared to females throughout all age ranges. Additionally this might also be due to the more prevalent risk variables for vascular dementia, like those of intense hypertension and coronary heart disease, in males. In total, 66% of people who are observed to be effected in dementia are women. The ratio differs with the age range, although: woman estimates for only 37% of individuals with dementia in the range of sixty five and sixty nine, yet 79% of individuals with dementia between ninety and overhead. An approximately 5.5 million adults in the United States are presently affected. In 2014, a number of 93,541 deaths from AD emerged in the United States at an age-adjusted rate of 25.4 deaths per 100,000 population (to the 2000 generic population) which is a boost of 54.5 percent particularly in comparison to the 1999 rate of 16.5 deaths per 100,000 population [1]. The number of symptomatic instances in the United States is forecast to an expansion rate of 13.2 million by 2050 without improvements in treatment [2]. In the United States, it is the sixth most important reason for death, accounting for 3.6% of all fatalities in 2014 [3]. Total and persisting healthcare costs in the United States for individuals with Alzheimer's disease and other neurological diseases are forecast to complete USD 259 billion in 2017, of which over two-thirds are anticipated to be financed by government sources such as Medicare and Medicaid [4]. Currently in Bangladesh, there are no accurate epidemiological data on AD. Some information on the amount of AD patients in Bangladesh are accessible [5]. As our country is a low middle-income nation it is not yet equipped to handle AD. The knowledge of AD is not satisfactory yet. Currently, most people in the nation are in the group of youthful. There will be enormous elderly individuals in this nation in less than 20–30 years. So, at that time there will be more chances of AD happening [6].

1.1 Organization of our paper

In this paper, we have discussed about the background information in chapter 2. The dataset description, proposed model, result analysis and discussion are described consecutively in chapter 3, 4 and 5. Lastly, chapter 6 consists of conclusion and future work.

1.2 Origin

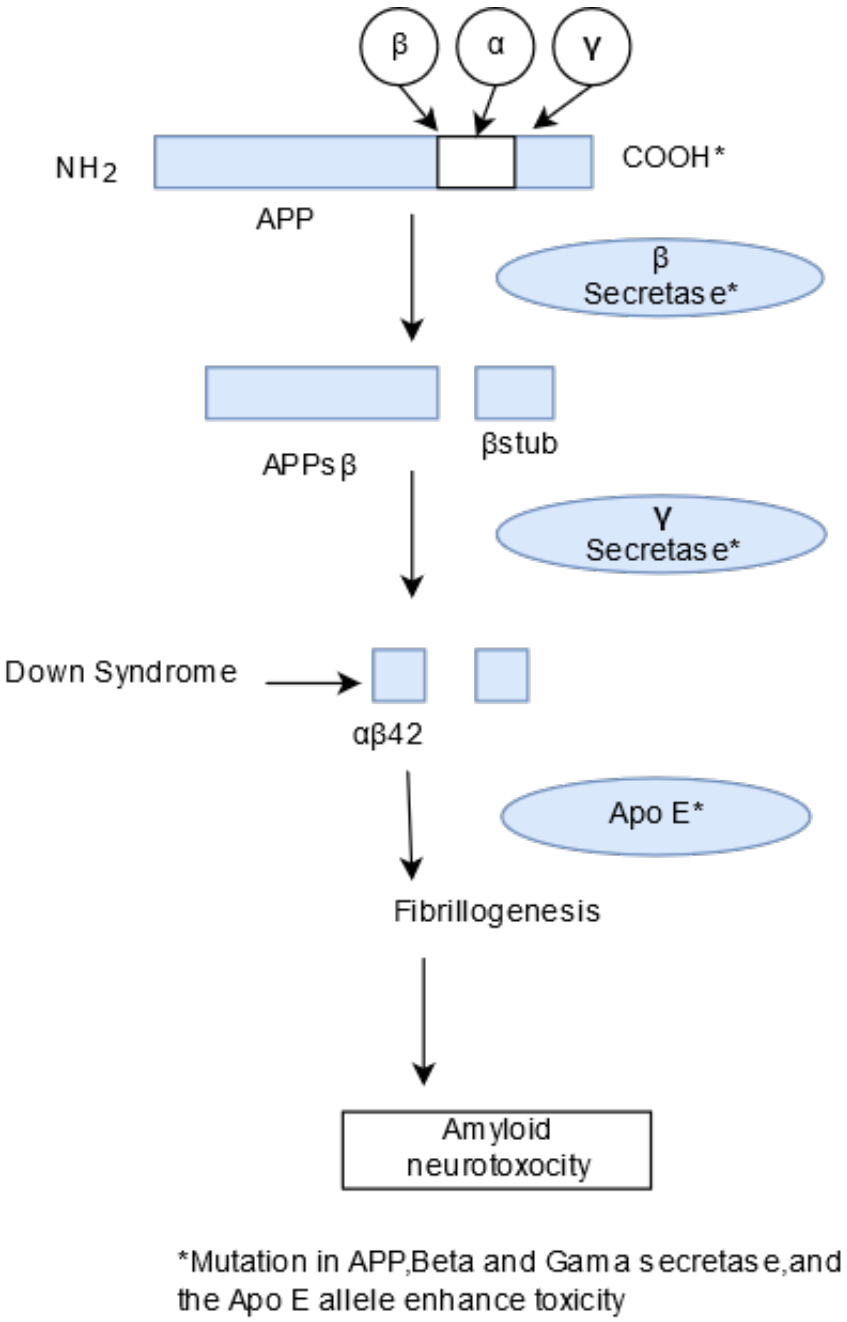

Figure 1.1: Amyloid neurotoxicity [5]

It is assumed that Alzheimer's disease will begin 20 years or more before symptoms develop, with minor brain modifications that are imperceptible to the person af-

fected [7].The disease usually begins in the 40's and 50's. Basically, these mutations involved early onset of AD, and only offer an explanation for a tiny proportion (less than 1%) of entire cases. There is a definite structure in some Alzheimer's early-onset disease (AD) patients. Commonly, personal memory impairment is the major and most notable feature of AD. Many patient have focal cortical, non-memory symptoms, such as linguistic problems, visuospatial, or exclusive features. These occurrences are connected to explicit patterns of atrophy and with a young age at onset. Age is not the defining factor of phenotype; underlying variables, especially genetic factors, also appear to affect the constitution and lead patients to age younger or older at the start. Identifying genetic variations associated with complicated diseases is considered one of human genome research's most notable. Apolipoprotein E (APOE) [8] (Figure 1.1) is an identified gene as the AlzGene database's most substantial risk factor for AD[9]. Patients with anomalous early-onset disease rarely carry allelomorphic APOE, which is the most significant risk problem in patients with AD to decrease their onset age. In addition, the genotype of APOE 4 seeks to incline patients in the medial temporal domains to vulnerability, which ends in a state of mind. On the contrary, APOE-negative patients are severely vulnerable to cerebral network weakness on the far side of the medial temporal lobes. The atrophy pattern is most probably concerned by different variables, but these are unknown at the moment. Some susceptibility genes are also being researched, of which apolipoprotein E gene polymorphisms have received the greatest attention, with earliest medical studies indicating that it is present in about 90% of early onset instances (which happen predominantly after 60 years of history and do not have an inherent autosomal dominant inheritance mode). Meta-analysis of recent epidemiological research has shown that while ApoE e4 is extra common in all forms of AD than in controls, it is especially related to late onset rather than early onset variant. ApoE e4 is therefore considered to be mostly associated with AD family cases of late onset. Having one copy of the ApoE e4 gene raises up to four times a person's risk of developing AD. Someone with two copies of ApoE e4, one from each parent, has an increased chance of 10 instances and in the past age of onset than individuals who inherited one e4 allele, but only about 2% of the population has two copies of e4. The gene's most common type is e3. Approximately 60% of the population has two copies of ApoE e3 and is at average threat, which, by the early 80s, will increase the disease by about half. Nearly one in six people has at least one copy of ApoE e2. This gene shape delays the start and reduces the risk of AD. The smallest risk is to individuals with two ApoE e2 copies. Understanding that this gene impacts risk is crucial and is no longer a predictor of whether or not someone is going to stimulate AD. Although ApoE e4 increases the danger of disorder development, it is no longer certain. Many individuals who reinforce AD do not have an ApoE e4 gene, and some individuals with e4 do not develop the disease anymore. It is now recognized that ApoE is not now the 'cause ' of AD, but alternately an significant link in an organic chain of occurrences, AD performing itself much less as a single disease mechanism and more the end result of the inability of multiple neuronal compensatory and repair processes to cope with a few aging-related aggressions. An interactive effect with ApoE in AD has now been confirmed in relation to a multitude of distinct opportunity variables so that ApoE e4 carriers may be more prone to a number of harmful environmental variables such as physical inactivity, saturated fat intake, consumption of alcohol, diabetes, increased BP and low B12/folate.

1.3 Motivation

The Alzheimer's Disease Research (ADR) program has been granting almost 120 million USD since 1985 to promote promising research in areas from molecular biology to genetics to epidemiology. Nearly 100 excellent biomedical research projects are presently being supported by ADR. Today, every 67 seconds, somebody develops AD in the U.S. By the year 2050, a unique AD patient is anticipated to evolve in almost every thirty three seconds, leading in almost one million unique instances over each year, and an approximate incidence of between 11 million and 16 million is anticipated. In 2013, 84,767 deaths from AD were reported in statutory death certificates. The actual amount of casualties to something that AD contributes (or deaths from AD) is estimated to be much higher than the number of fatalities from AD published on official documents [10]. In Bangladesh, Alzheimer disease is not as alarming as other nations especially United States, however, in the near 20 to 30 years our young generation would be elder which consists of a major part of our total population. Additionally, the number of smokers and people with obesity which contributes for causing Alzheimer is increasing. So, therefore, there is a huge probability that Alzheimer will be a threatening disease in the near future also in Bangladesh. The motivation behind choosing this topic as our research topic is to try to predetermine this disease as it a threatening and alarming disease throughout all over the world. Also distinguishing the features between Alzheimer's disease and MCI as it this difficult to differentiate the features of this two diseases. Currently, many researchers are working in this medical imaging field. It is a promising research field and there are chances of a lot of improvements and chances of making a tremendous breakthrough in predicting AD. The purpose of this research is predicting Alzheimer Disease earlier through sMRI image analysis by extracting features.

1.4 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 result with minimum effort.

Chapter 2

Background Information

2.1 Literature Review

In latest years, significant progress in neuroimaging has given possibilities for the control of neurological diseases, resulting in improved early and accurate identification of AD [11],[12]. In genetic research, machine learning models were used to explore genetic variants most associated with complex diseases [13]. 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 [14]. Additionally, magnetic resonance imaging provides a spatial resolution and smart distinction [15],[16],[17]. These developments have enormous potential for slow realization of the technology of medical imaging, analyzing medical data, medical diagnostics and general healthcare. Another significant region of implementation is sophisticated registration of deform able images, allowing quantitative analysis across distinct modalities of physical imaging and over time [18]. Many trials have therefore used structural MRI (sMRI-) mostly based on biomarkers to classify AD[19],[21],[22] explaining brain atrophy and altering the volume of brain tissues. Similarly, useful magnetic resonance imaging (fMRI) [23] is often used to characterize the hemodynamic response related to neural activity and functional or structural properties, which could be used to define medical specialty abnormalities in the entire brain at the point of connectivity [24],[25]. Promoters have increasingly introduced multi-national Alzheimer's disease (AD) research to enable registration and address local registration criteria. Geographic areas differ in many sizes, which may influence the development or assessment of disease. In that study they evaluated disease progression across geographical areas using placebo statistics from four big, multi-national clinical studies of research compounds designed to target AD pathophysiology , to help scientists construct and execute Phase 3 AD tests. In the result, Eastern Europe / Russia showed the largest physiological and operational decrease from baseline among all regional populations ; Japan, Asia, and/or Singapore. The least physiological and operational decrease was shown by America / Mexico [26]. It is a serious neurological brain disorder that can not be healed, but previous diagnosing it early can assist to treat adequately and prevent harm to brain tissue. Alzheimer's disease (AD) identification and classification is complicated because the signs separating Alzheimer's MRI samples can sometimes be found in normal healthy brain MRI samples of elderly people.In another research, a model for identifying and classify-

ing Alzheimer's disease using information analysis of brain MRI is discussed. They developed a deeply convolutional neural networks ensemble and clarified superior effectiveness in the Open Access Series of Imaging Studies (OASIS) dataset [27]. Hippocampal formation from pathological and MRI studies is known to be considerably atrophied in developed Alzheimer's disease. However, it is uncertain when the earliest changes occur in the HF. In some research, a longitudinal study of asymptomatic people at risk of autosomal dominant Alzheimer's family disease is conducted to evaluate presymptomatic changes in HF. Seven risky pedigree members of the Alzheimer family disease associated with the amyloid precursor protein 717 valine to glycine mutation underwent three-year serial MR scanning and neuropsychological assessments. These assessments were compared with 38 normal control results. During the research, three topics at risk were clinically affected. Volumetric HF measurement showed asymmetric atrophy formed before symptoms occurred in these topics. In conjunction with hippocampal loss, verbal and visual memory measurements decreased. A loss of up to 8 percent per annum of the HF quantity occurred in the 2 years during which symptoms first occurred. These results may have consequences for Alzheimer's disease early diagnosis [28]. In another study, MRI-based volumetric measurements of hippocampus, parahippocampal gyrus, and amygdala are performed in 126 cognitively normal older subjects and 94 patients with possible AD. The outcomes of the Clinical Dementia Rating (CDR) categorized AD diagnosis according to the criteria of NINDS / ADRDA. CDR 0.5 has been classified as very moderate, CDR 1 has been moderate, and CDR 2 has been classified as moderate disease severity. The amount of each measured MTL structure also reduced with age in patients with AD. The amount of each MTL structure in patients with AD was significantly smaller than control subjects (p is less than 0.001). Total hippocampal volumetric measurements were best used among the several MTL measures to discriminate control subjects from AD patients. The mean hippocampal volumes were as follows for AD patients: very small AD (CDR 0.5) -1.75 SD below the control mean, low AD (CDR 1) -1.99 SD and moderate AD (CDR 2) -2.22 SD [29]. A short overview is provided in a research of the recent innovations and associated machine learning problems appropriate for the evaluation of medical image and image analysis. A briefly overview of deep learning with reference points was given, showing how deep learning was applied in the general MRI processing chain, from procurement to image extraction, from clustering to prediction of disease. A starting point for people interested in researching and perhaps contributing to the field of medical imaging in deep learning was also created by highlighting powerful educational resources, state-of - the-art open source code and compelling sources of data and medical imaging issues [30]. Another research by van de Pol LA et al demonstrates that in frontotemporal lobar degeneration (FTLD) and its three clinical subtypes, using volumetry and a visual evaluation level to investigate hippocampal atrophy on MRI compared to Alzheimer's disease. In an experiment, 42 FTLD patients (17 frontotemporal dementia, 13 semantic dementia, and 12 progressive non-fluent aphasia), 103 patients with Alzheimer, and 73 controls were included. Hippocampal volumetry and the easily applicable medial temporal lobe atrophy (MTA) rating scale were used to evaluate hippocampal atrophy. Multivariate variance analysis for repeated measurements showed an impact on hippocampal size by the diagnostic group. An important side-by-side diagnosis (left v correct) interaction occurred. Frontotemporal dementia and semantic dementia

revealed paired hippocampal atrophy, and in semantic dementia the left hippocampus was smaller than in Alzheimer's disease. No significant hippocampal atrophy has been recognized in non-fluent progressive aphasia.

2.2 Convolutional Neural Network

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. A image which is in digital form is a discrete portrayal of visual information. CNN is introduced a potent system class for challenges with image recognition [31]. They are a influential method that can operate on the random inputs straight [32]. It includes a grid-like pattern of pixels that includes pixel values to indicate how luminous every pixel must be and what pigment it must be. The basic feed forward neural networks described above can be used in principle when it is needed to use neural network in images. However, having connections from all nodes in one layer to all nodes in the next layer is extremely inefficient. A cautious reducing of the links based on domain expertise, i.e. the image composition, results in much better performance [33].CNNs allow data-driven, extremely representative, hierarchical image functions to be learned from required training data.Currently, There are three main methods that use CNNs adequately to distinguish medical images: train the CNN from scrap, use them CNN attributes off-shelf, and monitor unsupervised pre-training with supervised adjustment. There is other efficient technique is to transfer learning to medical image functions, i.e. pre-trained CNN models from the existing image dataset for adjustments [34].Examples emphasize and motivate the importance of many independent input signal arrays, adaptive templates, and multi-layer capability. It shows how a partial differential equation can be generated simply as a wave-type [35].

2.2.1 Convolution Layers

Convolution is the first layer to extract characteristics in an input image. Convolution keeps the link between pixels by using small squares of input data to know image feature (Figure 2.1). It is a mathematical operation requiring two inputs such as the image matrix and a filter or kernel. The prior layers ' activations are transformed into a number of small parameterized filters, often in the convolution layers of size 3 x 3. One produces a rapid decrease in the number of weights to be learned by having each filter share exactly the same weights across the whole input domain, i.e. translation equivariance at each layer. The intention for this sharing of weight is that characteristics are probable to appear in one portion of the image in other areas as well [36]. Let us take an example of a 5x 5 with 0, 1 and 3x 3 filter matrix image pixel values. Then the convolution of the 5x 5 image matrix multiplies by the matrix of 3x 3 filters called "Feature Map." Convolution of an image with distinct filters by applying filters can conduct activities such as edge detection, blurring and sharpening.

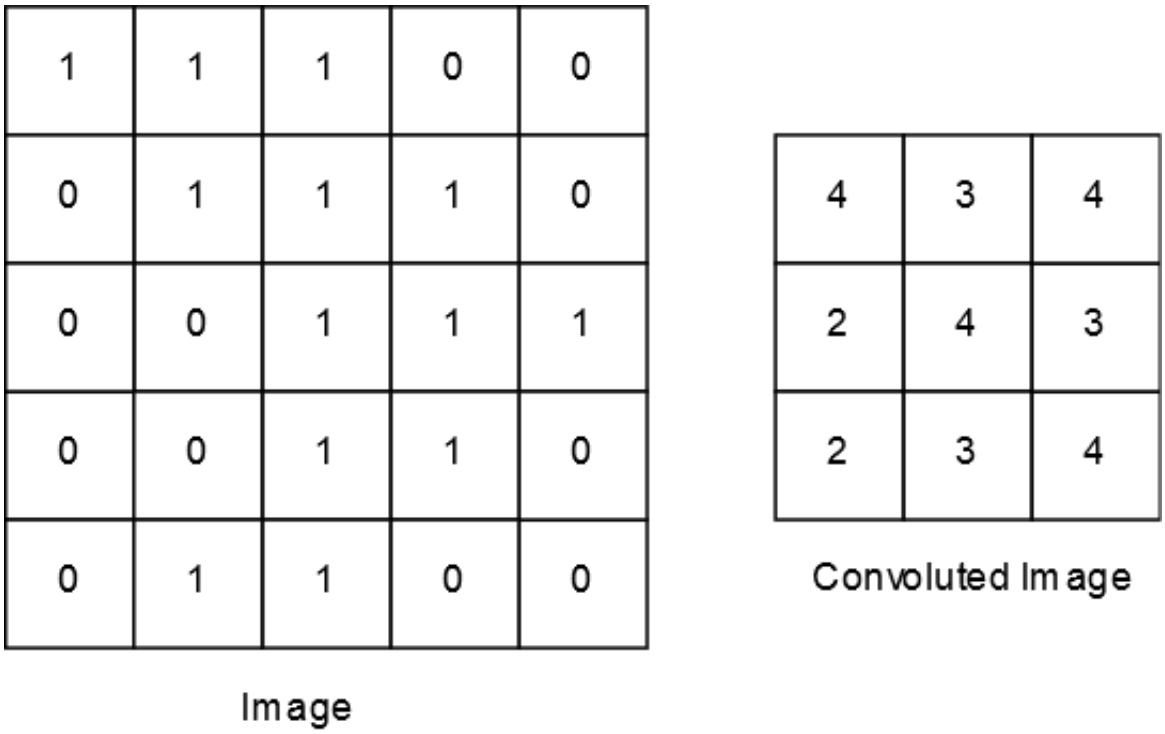

Figure 2.1: Convolution Layer [40]

2.2.2 Activation Functions

Feature maps are supplied through nonlinear activation functions from a convolution layer. Activation functions are crucial for a Convolutional Neural Network to understand something really difficult and inconvenient functional mapping among inputs and variable and make logical sense. It is shown that the universal estimation property still validates a neural network with unrestrained activation functions [37]. This enables almost any nonlinear function to be approximated throughout the neural network [38],[39].Some well known activation functions are Step function, Linear function, Sigmoid Function (Figure 2.3), Tanh Function, ReLu (Figure 2.2). ReLu is cheaper than tanh and sigmoid computationally because it includes easier mathematical operations. While designing deep neural networks, that's a fine point to consider.

2.2.3 Pooling

The pooling layer is responsible for lessening the spatial magnitude of the Convoluted Feature. This reduces the information processing computing capacity by decreasing the dimensions. Taking small grid areas as input, pooling activities (Figure 2.4) create single figures for each area. Usually the max function (max-pooling) or the average function (average pooling) calculates the amount [40]. On several image recognition benchmarks, max-pooling is simply substituted by a convolutional layer with enhanced step in the absence of dropping of precision [41]. It is effective for extracting rotational and positional invariant dominant features, thus preserving the model's effective training process.

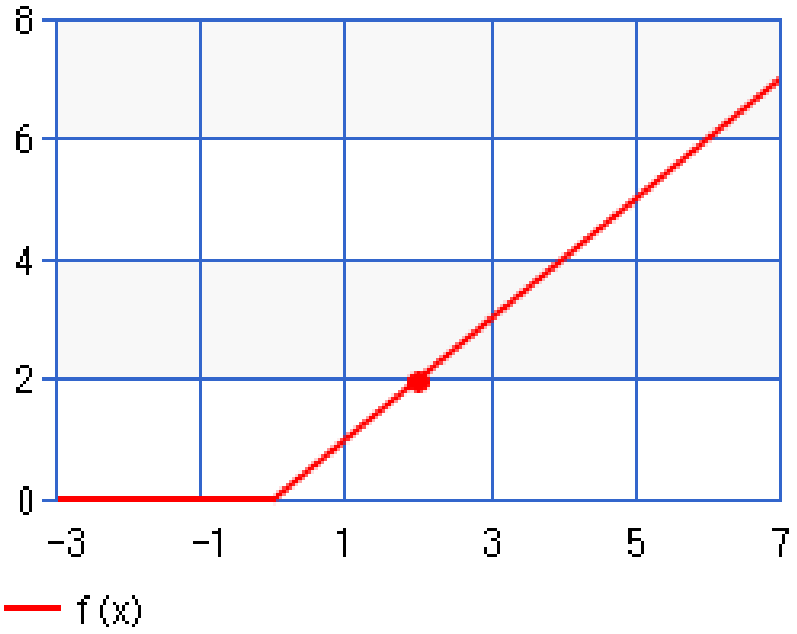

Figure 2.2: Rectified Linear Unit(ReLu) Function [40]

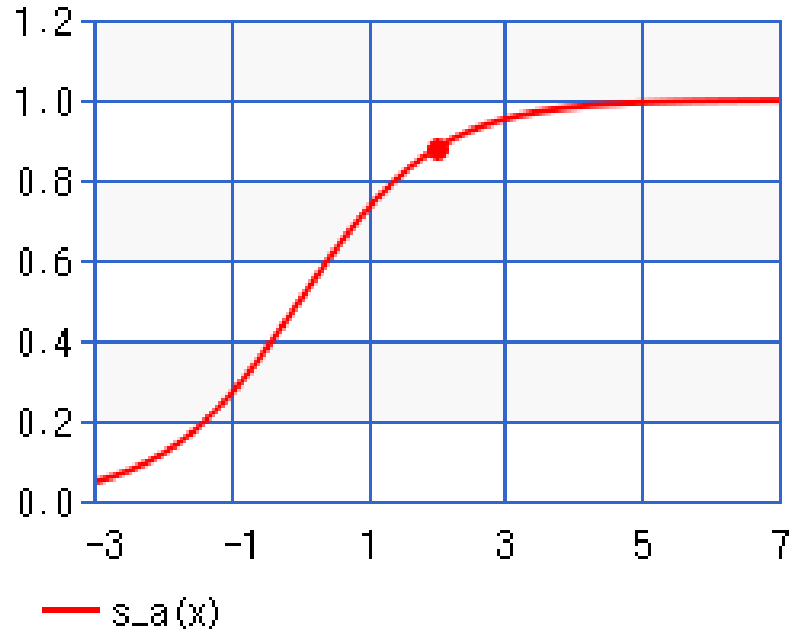

Figure 2.3: Sigmoid Function [37]

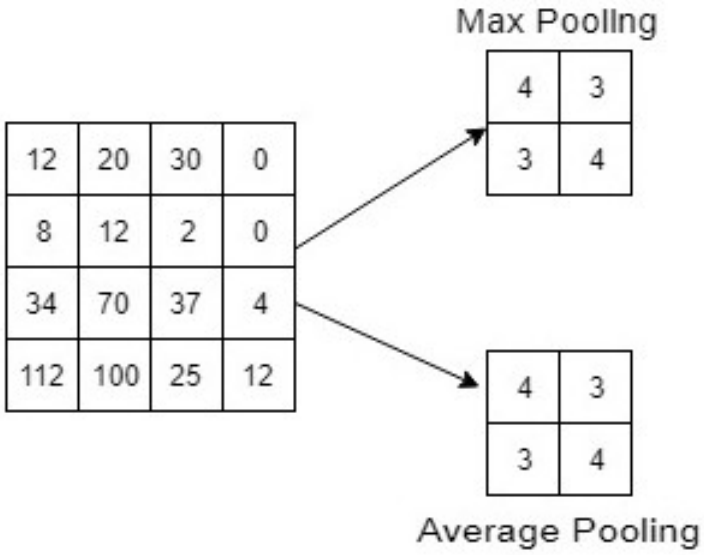

Figure 2.4: Types of Pooling layer [40]

2.2.4 Dropout

Dropout (Figure 2.5), the most effective neural network regularization method. Dropout and other feature noise systems regulate overfitting by corrupting the training data artificially. Dropout holds out a type of adaptive regularization for generalized linear models [42]. By averaging distinct models in an assembly, one tends to perform better than single models. Dropout is an average stochastic selection method based on the neural network [43]. By randomly removing neurons during training, one ends up using slightly distinct networks for each training data batch, and the weights of the trained network are tuned based on optimizing different network versions. Some layer inputs are randomly overlooked or dropped out during training. This has the impact of making the layer look and be treated as a layer with a distinct amount of nodes and connectivity to the previous layer. In effect, a distinct view of the configured layer is conducted with each update to a layer during training.

2.2.5 Batch Normalization

By modifying and optimising the activations, we normalize the input layer. For instance, if we have features from 0 to 1 and some from 1 to 1000, we must standardize them to accelerate training. If this privileges the layer of input, it should be the same for the values in the hidden layers that are constantly evolving and get ten or more refinements in the rate of training. It (Figure 2.6) decreases the quantity by shifting across the hidden unit values (covariance shift). Batch normalization legitimizes the output of a preceding activation layer by removing the interval mean and splitting it by the standard batch interval in order to boost the sustainability of a neural network [44].

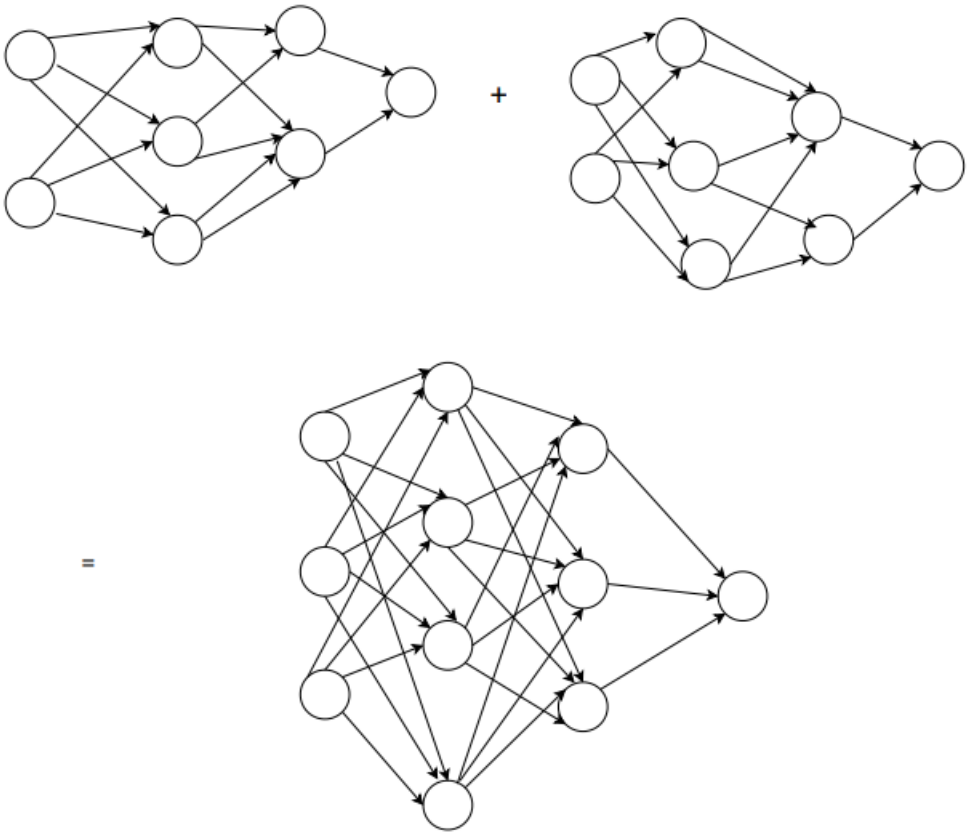

Figure 2.5: Dropout in Convolutional Neural Network [40]

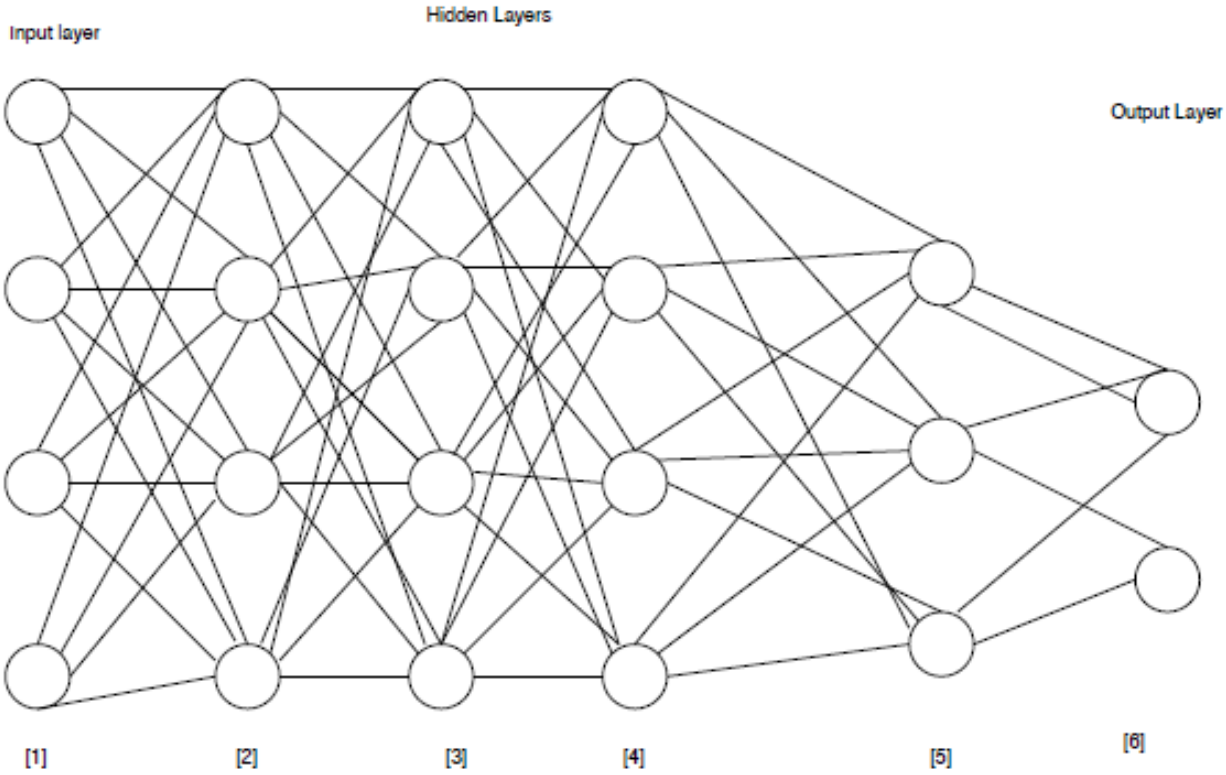

Figure 2.6: Batch Normalization in Convolutional Neural Network [44]

2.3 Transfer Learning

Transfer learning is a machine learning method in which a model set up for a task is utilized as the entry point for a second task model. It is a prevalent deep learning approach where pre-trained models are used as the point of departure for computer vision and natural language processing tasks owing to the large computational and time resources required to build neural network models on these subjects and the immense ability leaps they provide on related problems. It is possible to use a new dataset smaller than the original dataset used to train the transfer learning of the pre-trained model. Traditional machine learning makes a basic assumption: the training and test data should be dispersed in the same way. Labeling the latest information can be expensive and throwing away all the old information would also be a waste [45].It is frequent to transfer learning with predictive modeling problems that use image data as input. This can be a predictive job that needs input images or video data. A comprehensive pre-trained teaching model for a large and challenging task of image classification such as the ImageNet 1000-class picture classification competition is frequently used for such problems. Unlabeled information is much easier to achieve than standard semi-supervised learning settings and makes self-educated learning frequently appropriate to many realistic learning problems [46]. Transfer learning highlights the problem of how to use lots of labeled information in a input domain to address associated but different problems in a target domain, even when there are different variations or features in practice and testing. Methods based on features were commonly used in many fields linked to transfer learning . Domain-specific data in associated assignments is used in multi-task learning to train various classifiers together so that they can benefit one another [47]. There are many transfer learning techniques, including freezing layers and training on subsequent layers, and using a small level of learning. Some of this optimization is often accomplished by trial and error, so there are options to experiment with various alternatives.

2.4 Neural Network Models

The organizers of the Neural Information Processing System (NIPS) conference even figured out in 2000, after becoming highly popular back in 1990's, that the phrase "neural networks" in the title that was proposed inversely corresponded with approval.On the other hand, there were favorable correlations with support vector machines (SVMs), Bayesian networks, and variational methods [48]. Neural networks are a sequence of algorithms intended to acknowledge patterns that are narrowly designed after the human nervous system. They evaluate sensory information by perceiving, listing or grouping the raw data input. The patterns that they acknowledge are numerical, encapsulated in vectors that need to be translated into all real-world information, be it images, sound, text or time trilogy. It can also extract features for clustering and classification that are supplied to other algorithms. It assists us to classify and cluster. They assist group unlabeled information by similarities between the instance inputs and classify information when they have a marked dataset to train. The precise structure of a gradient-following learning algorithm for fully recurring networks functioning in continuously analyzed time is measured and used as the baseline for realistic algorithms for temporally supervised

learning assignments. Such algorithms have the benefit of not requiring an accurately specified training interval to proceed whereas the network is working . On the other hand the problem is requiring non-local communication in the network being trained and being quantitatively costly. These algorithms enable networks with recurrent relationships to learn complicated processes that require whether defined or indefinite length retention of details over epochs [49]. Using minor adjustments in inner parameters, artificial neural networks were interpreted only as a machine that learns [50]. The neural network algorithms we used for our model are described below.

2.4.1 VGG19

VGG Net is created by the University of Oxford's VGG (Visual Geometry Group). VGG-19 (Figure 2.7) is a convolutional neural network based on more than one million pictures from ImageNet [51],[52].The "19" in VGG stands for the weight layers number in the network. VGG network is the idea of much deeper network with much smaller filters. It has increased number of layers, now it is a model with 16 to 19 layers. As the filter is very small, only three by three convolution layer with the periodic pooling all the way which is basically the smallest filter size. It is a convenient, elegant network architecture that can get 7.3 percent of the top 5 error on the challange of ImageNet. The stack of three by three convolutionary layers with stride 1 has the same likely to respond field as one seven by seven convolutionary layers but deeper and has more non-linearities and fewer parameters as well. The deeper the layers in this situation ; complete amount of layers with easy to train weights. Each filter is a set of weight looking at input depth and then produces one feature map, one activation map of all the responses of the different spatial locations.Then we can have as many filters as we want and each of them is going to produce a feature map. So, it is just like each filter corresponds to a different pattern we are looking for in the input that we convolve around and see the responses everywhere in the input create a map of these and another filter will convolve over the image and create another map. It is not needed to store all of the memory but also it calculates backward propagation so it needs a lot of those activation. That is why a large part of these has to be kept. According to ILSVRC'14 VGG is first in localization, second in classification. It has no Local Response Normalisation (LRN) and uses ensembles for best results.FC7 which is basically a layer before the classification layer features generalize well to other tasks. VGG19 is basically a pre-trained model for keras. The pre-trained model is the process where smaller number of networks is converged to be used as initializations for the larger network. It is used in deep learning image classification problems. The network is deep in 19 layers and can classify pictures into 1000 groups of objects, including keyboard, mouse, pencil and many livestock. As a consequence, the network has learned from a broad spectrum of images wealthy feature depictions [51],[53]. VGG-19 model utilizes Stride and Pad of 1 and 3x3 filters, together with 2x2 Stride 2 max-pooling layers. The VGG-19 is a broader CNN with more layers than AlexNet. In all convolution layers, it uses small 3x3 filters to reduce the quantity of parameters in such deep networks and is best used with its error rate of 7.3%. It is undoubtedly one of the most important articles because it strengthened the concept that CNNs must have a deep layer network in order to work with this hierarchical depiction of

Figure 2.7: VGG19 Model[51]

visual data [53].

2.4.2 Inception V3

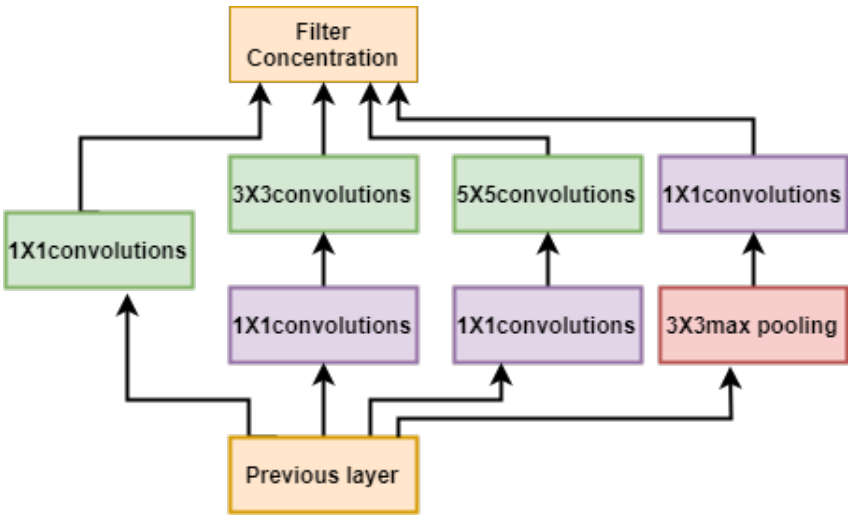

Figure 2.8: Inception Model [55]

While VGG achieves remarkable precision on the ImageNet dataset, its implementation on anything but the smallest GPUs is a issue owing to enormous memory and time requirements. Because of the big range of convolution layers, it becomes inefficient. GoogLeNet developed a module called the Inception (Figure 2.8) 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 is 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. There is only one back propagation all the way through, like an addition at the end in case of drawing up a computational graph, a final signal is got, the gradients are taken and back plotted all the way through as if they are added together at the end in a computational graph. Last few layers of the network is a fully connected layer preceded by a softmax layer to try to make a prediction. e branches take some hidden layers and it tries to use that to make a prediction. Then there is other side branches that take hidden layers, passes it through to a few layers like a fully connected layer and then it has the softmax try to anticipate the output label. Another detail of this network is that it facilitates to fortify that the features even in the hidden units are calculated, even in the intermediate layers that they are not too big of a deal to anticipate the output of an image, and this attempt disappears to have a normalizing impact on the network and how this network can be avoided from overfitting. The inception network was developed by authors at Google who called it

GoogleNet. This decreases the complete amount of parameters dramatically. Using a large network width and depth allows GoogLeNet to remove the layers of the FC without affecting precision. It achieves top-5 precision of 93.3 percent and is much quicker than VGG while working with ImageNet.

2.4.3 ResNet50

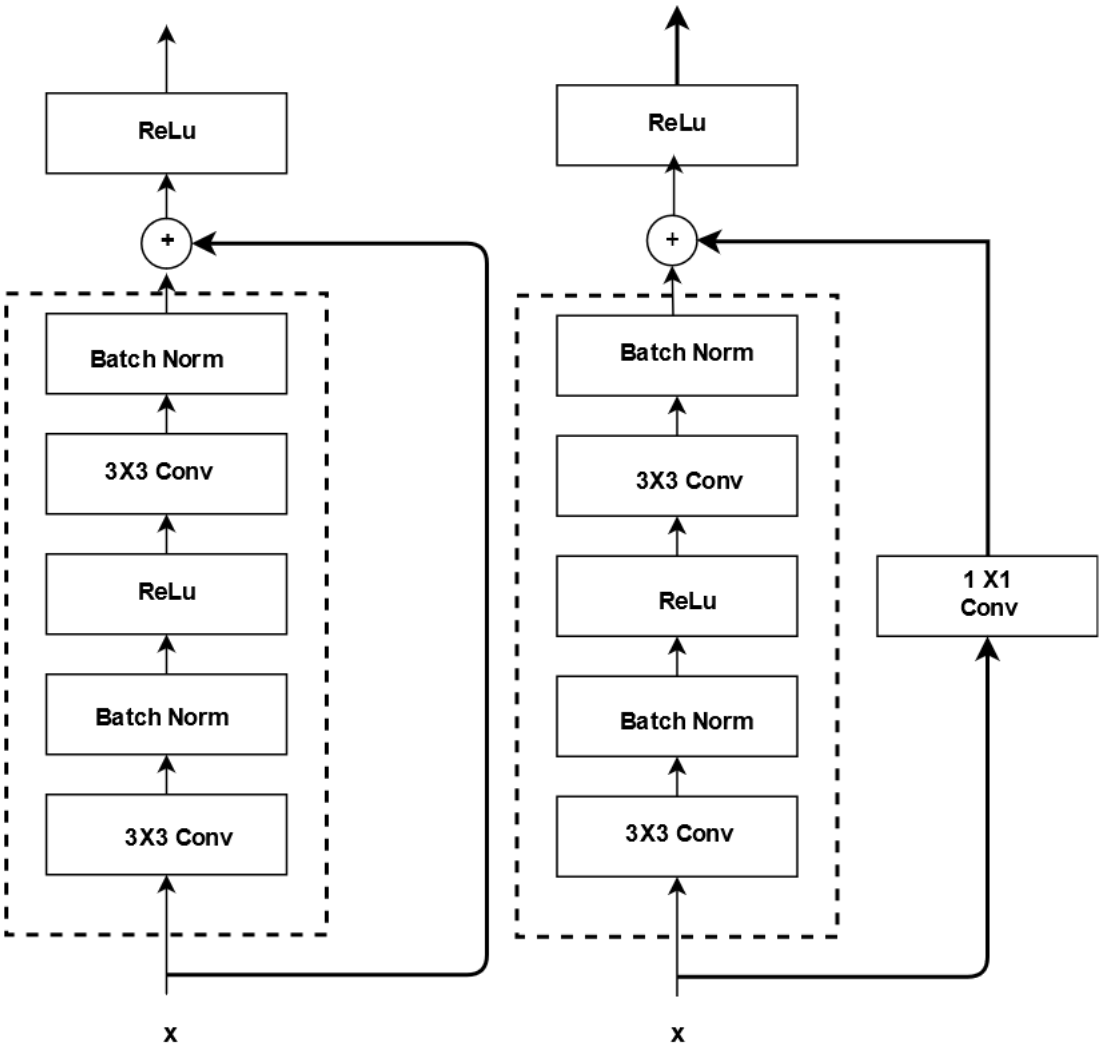

Figure 2.9: Residual Network Model [56]

A degradation issue has been revealed when deeper networks begin to converge: with the network depth increasing, precision becomes saturated and then quickly degrades. Recently, residual networks (ResNets) have accomplished state-of - the-art computational vision challenges. ResNet (Figure 2.9) has very deep networks, much deeper than any other networks using residual connections. They had 152 layers model for ImageNet.They were able to get 3.5 of 7 percent top of the error. The deeper designs perform worse, but overfitting does not cause them. ResNet creators ' hypothesis is deeper model is more difficult to optimize. At least as well as the steeper model should be able to perform the deeper model. Duplicating the learned layers from the shallower models and setting extra layers for mapping iden-

tification is a solution by development. 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". This model was ImageNet's 2015 challenge winner. He et al. states in their 2015 paper, the ResNet architecture has become a seminal work, showing that immensely deep networks can be trained with the help of using conventional SGD (and a rational initialization function) by using residual modules. ResNet's basic discovery has enabled us to effectively train incredibly profound neural networks with 150+layers. Because of the issue of disappearing gradients, it was hard to train very profound neural networks before ResNet.The notion of skip connection was first launched by ResNet.The diagram below shows the link to skip. The figure on the left is stacking layers of convolution one by one. We are still stacking convolution layers on the right as before, but we are now adding the initial input to the convolution block output as well.This is called the link to skip. Since X, above are two matrixes, X shortcut can only be added if they have the same shape. So if the normal convolution + batch operations are carried out in such a way that the output shape is the same, then we can simply add them. Even though ResNet is much deeper than VGG16 and VGG19, the model size is significantly smaller due to the use of global average pooling rather than full-connected layers — it thus minimizes the model size for ResNet50 to 102 MB.

2.5 Structural Magnetic Resonance Imaging (sMRI)

Magnetic resonance imaging (MRI) is a type of testing that uses strong magnetic fields and radio waves to produce extensive images of our body's inner part.MRIs are often divided into structural MRIs and functional MRIs (fMRIs). Whereas structure and function are often intrinsically intertwined within the brain, it is difficult to establish the division between structural and functional imaging. Functional imagery definitions are diverse and mostly broad6 and will always be to some extent arbitrary [54]. Structural magnetic resonance imaging (MRI) is a non-invasive technique for analyzing brain anatomy and pathology (as opposed to using functional magnetic resonance imaging (fMRI) to investigate brain function). This creates models that can be used for both clinical radiological reporting and thorough analysis. It is common for structural magnetic resonance imaging (MRI) using regular 1.5 T MRI unless the patient has some difficulties. Old patients may experience cortical atrophy and/or ventricular dilation associated with age. In both substantia nigra pars reticulate (SNR) and red nucleus, efforts were produced to identify quantitative distinctions between patients with PD and healthy controls using strong-resolution 1.5 T MRI. Structural magnetic resonance imaging (sMRI) (Figure 2.10,2.11) basically combines local variations in water content into distinct colors of gray to describe the shapes and sizes of the distinct subregions of the brain (Nora D. Volkow, ... Ruben Baler et al., 2014). Numerous sMRI studies have reported that addictive drugs in the prefrontal cortex (PFC), a brain region that promotes logical thinking, goal-oriented behaviors, planning, and self-control, can cause volume and tissue structure alterations. These modifications are likely to be linked to the cognitive and decision-making defects of drug addicts. In association with this assessment, another sMRI research discovered that people with a history of various narcotics abuse have lower prefrontal lobes than controls paired (Liu et al., 1998).

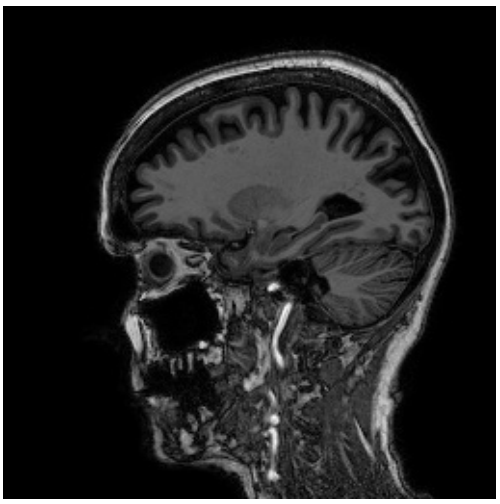

Figure 2.10: structural MRI image of Alzheimer patient (from the dataset ADNI1-Annual-2-Yr-3T-8-03 2019)

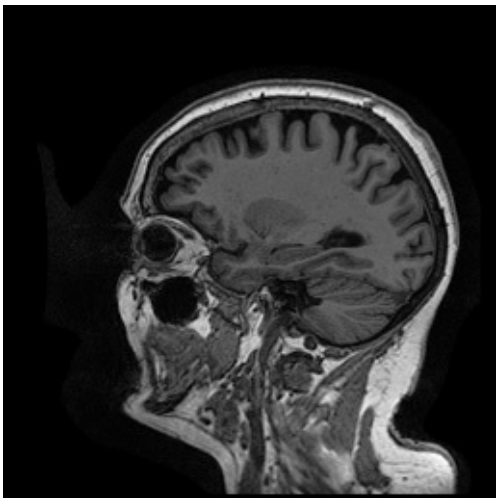

Figure 2.11: structural MRI image of Cognitively Normal (from the dataset ADNI1-Annual-2-Yr-3T-8-03 2019)

Chapter 3

Dataset Description

3.1 Introduction

In our research we collected 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, cognitive tests, CSF and blood biomarkers. They provides 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.

3.2 Dataset Details

We dealt with dataset from ADNI. The dataset is basically Structural MRI images.We used the recently uploaded dataset "ADNI1-Annual-2-Yr-3T-8-03 2019". which means ADNI server 1 collected data from people with 2 years of Mild Cognitive Impairment and Alzheimer (Table 3.1). The metadata was in CSV format which contained Image Data, ID, group, sex, age, visit, modality, description, type, acquired date and format. Tesla stands for the "T" in 3T. Tesla is described as the measuring unit used to describe the magnet force used in an MRI. Magnetic Resonance Imaging (MRI) is the magnet. This magnet is the fundamental basis for the acquisition of images in MRI. The power of the magnet directly impacts the quality of these pictures, but there are several other variables that evaluate which magnetic resistance is ideally suited for the individual being pictured and scanned or pictured for the particular portion of the body. The dataset was uploaded in 8 March, 2019. From the dataset we got 24915 female MRI 2D segmented slices from 3D images. Again, 25245 male MRI 2D segmented slices from 3D images. We trained our model with 4227 Cognitively Normal (CN) and tested it with 1057 Cognitively Normal. Also, trained with 2247 Alzheimer images and tested with 561 of it. Again, trained with 4718 Mild Cognitive Impairment images and tested with 1180 of it.

Table 3.1: Table of dataset

Image data	ID	Group	Sex	Age	Visit	Modality	Description	Type
40378	136_S_0426	AD	M	80	2	MRI	MPR;;N3;Scaled	NIFTI
120441	136_S_0426	AD	M	82	6	MRI	MPR;;N3;Scaled	NIFTI
119731	136_S_0426	AD	M	80	2	MRI	MPR;;N3;Scaled_2	NIFTI
66787	136_S_0426	AD	M	81	4	MRI	MPR;;N3;Scaled	NIFTI
119729	136_S_0300	AD	M	57	2	MRI	MPR-R;;N3;Scaled_2	NIFTI
66768	136_S_0300	AD	M	58	4	MRI	MPR;;N3;Scaled	NIFTI
40352	136_S_0300	AD	M	57	2	MRI	MPR-R;;N3;Scaled	NIFTI
120436	136_S_0300	AD	M	59	6	MRI	MPR;;N3;Scaled	NIFTI
119725	136_S_0196	CN	F	78	2	MRI	MPR;;N3;Scaled_2	NIFTI
66740	136_S_0196	CN	F	79	4	MRI	MPR;;N3;Scaled	NIFTI
40269	136_S_0196	CN	F	78	2	MRI	MPR;;N3;Scaled	NIFTI
120426	136_S_0196	CN	F	80	6	MRI	MPR;;N3;Scaled	NIFTI
40254	136_S_0195	MCI	M	80	2	MRI	MPR-R;;N3;Scaled	NIFTI
120423	136_S_0195	MCI	M	82	6	MRI	MPR-R;;N3;Scaled	NIFTI
81981	139_S_0195	MCI	M	81	4	MRI	MPR;;N3;Scaled	NIFTI
119721	136-S_0195	MCI	M	80	2	MRI	MPR;;N3;Scaled_2	NIFTI
69136	136_S_0184	CN	F	79	4	MRI	MPR-R;;N3;Scaled	NIFTI

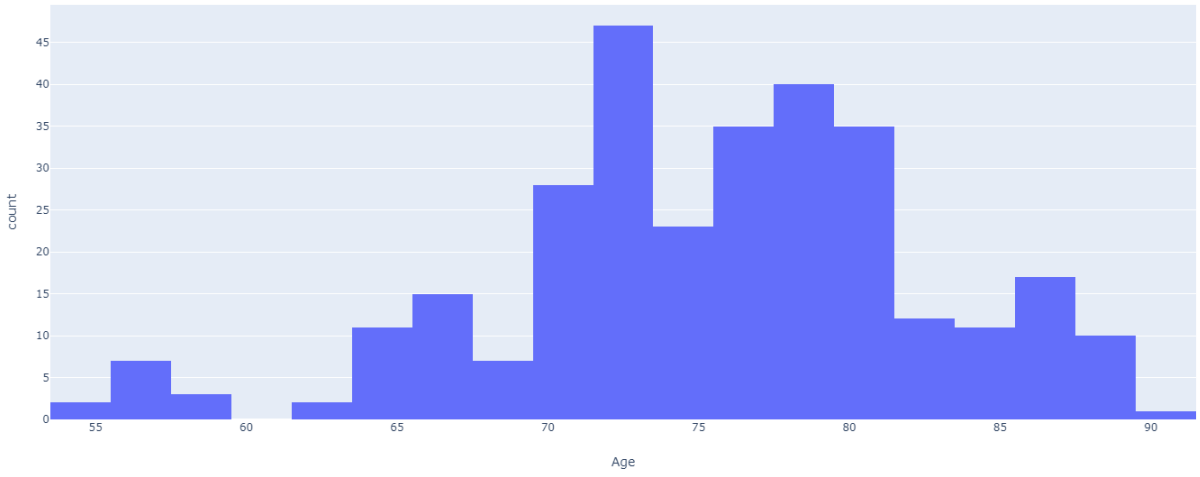

Figure 3.1: Age Diversity

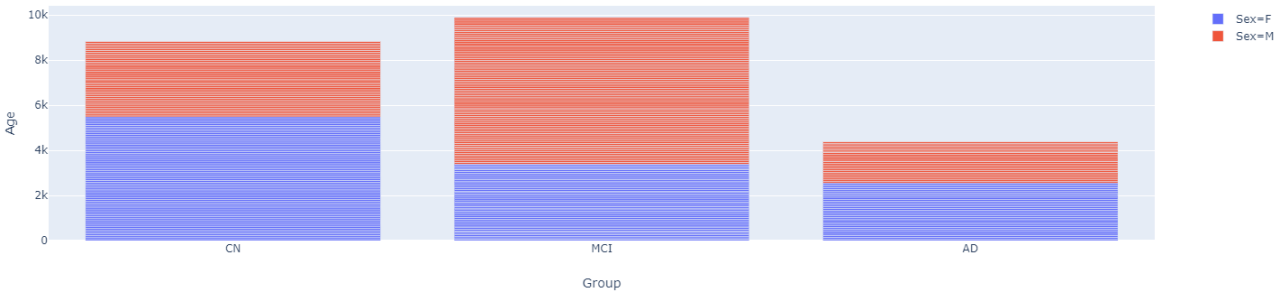

Figure 3.2: Age Vs Group and Sex

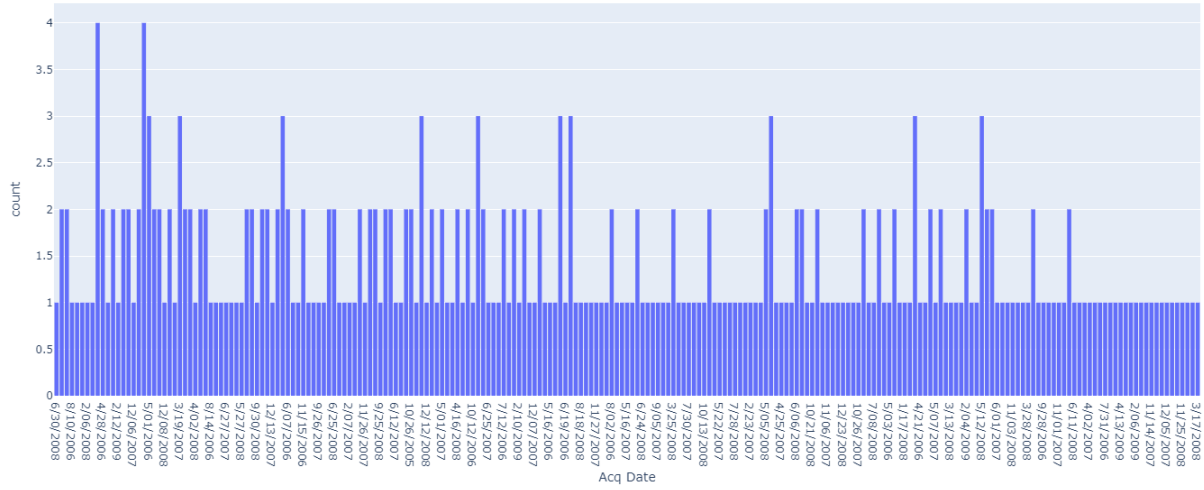

Figure 3.3: First Diagnosed Vs Visit Count

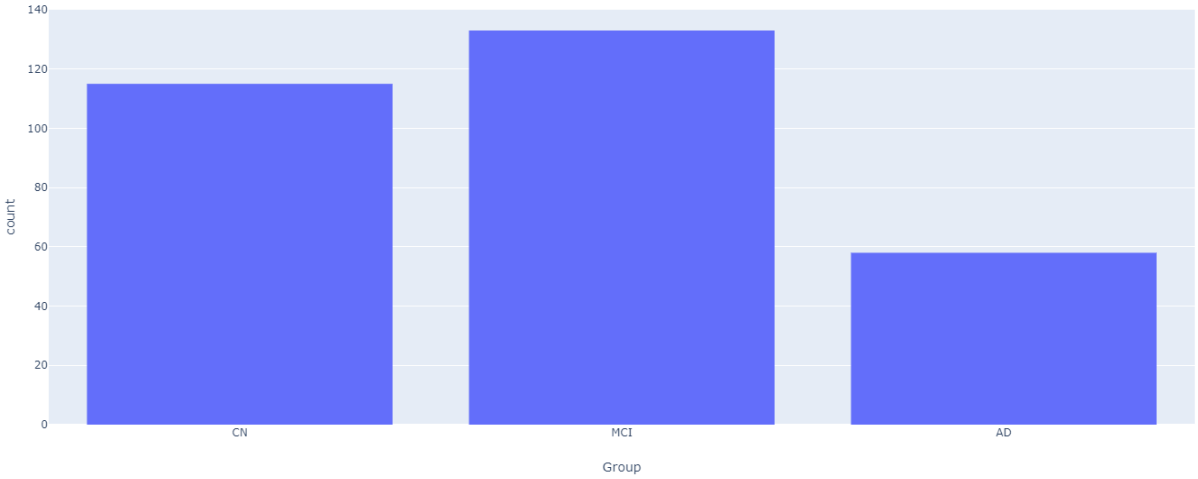

Figure 3.4: Group Count

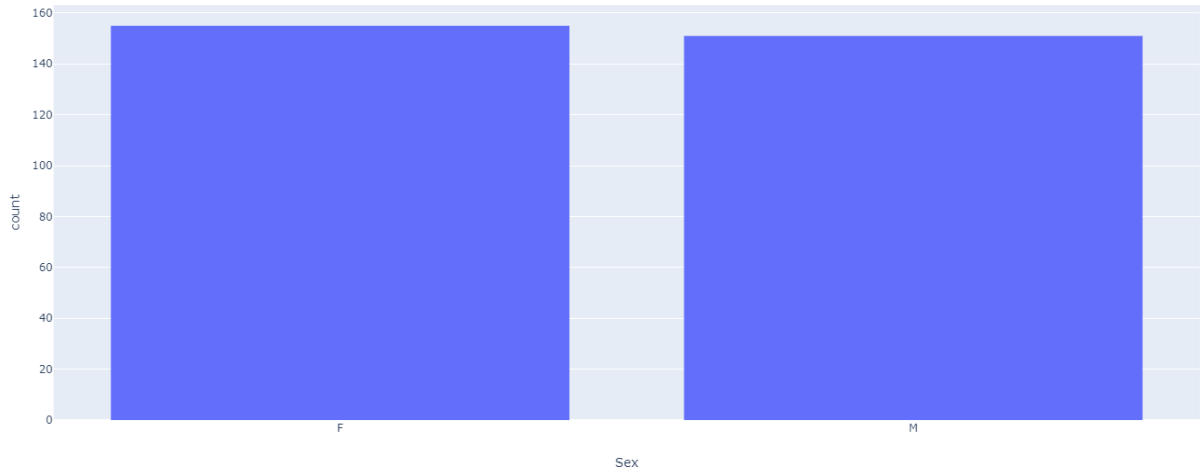

Figure 3.5: Sex Distribution

3.3 Dataset Acquisition Process

For acquisition of data, we visited Alzheimer's Disease Neuroimaging Initiative (ADNI)'s web server. This ADNI requires access permission to download any kind of dataset. After entering the web server we searched for modality: MRI, Field strength: 3 tesla, slice thickness: 2mm, weighting: T1, project: ADNI. The image type we searched for is original under search section: subject, study, visit, image. Display options: subject specific ID. For diagnostic and therapeutic purposes, medical imaging can be used, With regard to diagnosis, popular kinds of imaging that are inclusive of: CT (Computer Tomography), MRI (Magnetic Resonance Imaging), Ultrasound, X-ray. They each operate a bit separately to generate pictures of what is happening within the body. Within these, we used sMRI. The 3 T scanner's advantages are not limited to Magnetic Resonance Imaging. The 3 T scanner's enhanced spatial resolution enables high-quality vascular imaging. Thus, the need for invasive interventional catheter research can often be supplanted by 3 Tesla MR Angiogram research.

3.4 Dataset Processing Method

After downloading the data, we started to process the data. Data processing is essential because productivity and profits are increased, better decisions are being made, more precise and more consistent. Other benefits include further cost reduction, convenient storage, distribution and reporting, followed by better assessment and evaluation. The processing method is described below.

3.4.1 Catagorize 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 pyhton code to execute this process.

3.4.2 Segmentation of 3D files

We used Python package med2image version 1.1.2. Med2image is a convenient Python3 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 file and organized those according to user specific ID.

3.4.3 Extracting and attaching the image extension

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

3.4.4 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 from our study it is found that there are huge similarities between the features of AD and MCI MRI images. The Congnitively Normal class is created for comparing the normal subjects with the disease effected subjects. We divided the subjects into those groups against their user ID.

3.4.5 Train-Test split

The train and test part is divided into 80 percent and 20 percent accordingly. We distributed this in uniformed process. Because if we just splited them into random 80 percent and 20 percent then the distribution would not be uniformed. To overcome this, we selected 5 subjects, from the chosen subjects, 4 is taken to the training set and the rest one is given to the test set. That is how, the subjects were splited into train and test set.

Chapter 4

Proposed Model

4.1 Introduction

In this chapter, we will be describing the overall process model and methodology of our system that we proposed in a brief manner.

4.2 Structural MRI data acquisition

This applied to the classification of both binary and multi-class images. We had a folder with all the pictures we wanted to train your model on. Then, we also needed the 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 dataset from Alzheimer's Disease Neuroimaging Initiative program which is a organization to sharing Alzheimer's research data with the world.

4.3 Data Pre-processing

The preprocessing module involves standardization of inter-slice intensity, reduction of noise, and correction of non-uniformity intensity. First, all the images are loaded 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 the unreadable images and the images that are not useful for example, parts of image slices of neck portion that are not included in brain images. We filtered for the data has most recent acquisition date along with a filter consist of 3 Tesla magnetic resonance and slice thickness is 2mm. The MRI image we collected was structural magnetic resonance image and format was in NIfTI. NIfTI is an analyze-style data format which is volume of 2D slices. We organize and grouped the data by the specific patient identification. After doing so we had one class of 3 grouped data which were cognitively normal, Alzheimer disease and mild cognitive impairment. Hence one class is inefficient to convert a MRI NIfTI image 2D volume to 2D image slice.

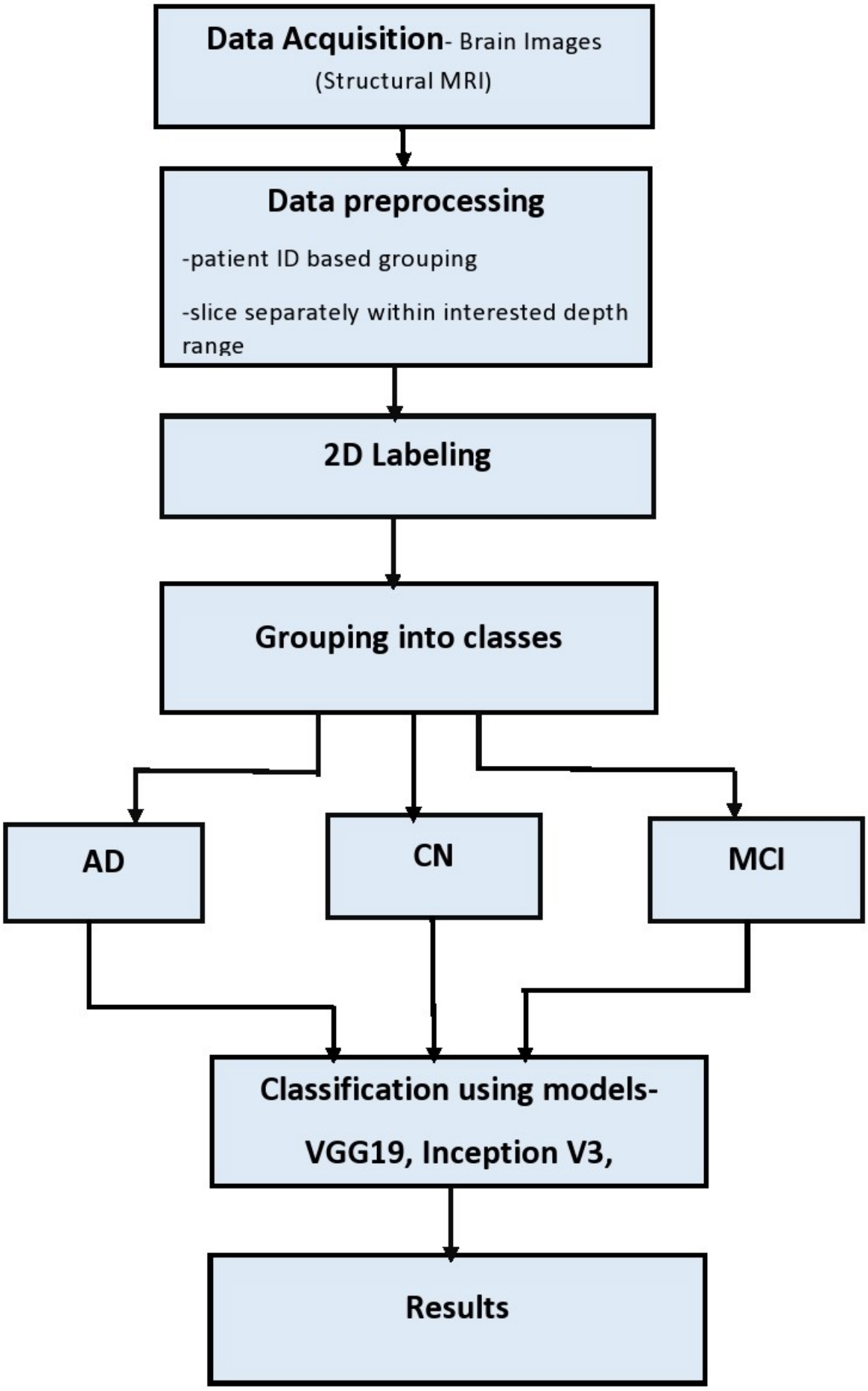

Figure 4.1: Block diagram of the proposed model

4.4 Grouping into classes

Therefore we again in a programmed way divided one class into 3 class for CN, AD and MCI. Then we used a medical image processing technique to slice those 3 classified set. The technique involved med2image, a python package, can slice medical scans into 2d image array. We got 180 image slices from one NIfTI file. From that 180 images per patient ID we dropped 20 slices from top and 20 slices from bottom. That remaining 140 images are our interested depth range. Going forward with the process we generalised all 3 classes to make 3 class unanimous. In this process we sequentially tagged each images object and separated those from image extension. After completion we reattached the extension. Completing this all process stated above has given us three unanimous group 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 do not represent the region we are featuring of a human brain we dumped that image and moved to the next one. After this crude process we had 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 CNN model using transfer learning. Therefore we split those 3 class each into 2 class as training set and testing set. The split was 80% and 20%. A simple python code used mod function to perform this split.

4.5 Classification Using 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 as well as the pre-trained weights. While the architectures are being trained on 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 network. 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 breakthrough 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 was calculated for the ImageNet dataset. The target was to achieve convergence as fast as possible. For out experiment, we

used a PC with 3.6 giga hertz i7 processor, 8gb ram and GTX 1080TI GPU. WE used mini batch gradient descent and a batch size of 20 was used. We reached the convergence (our target result) just after 15 epochs because of transfer learning. After that, the results of all the algorithms were compared and discussed.

Chapter 5

Result Analysis and Discussion

5.0.1 Results Analysis

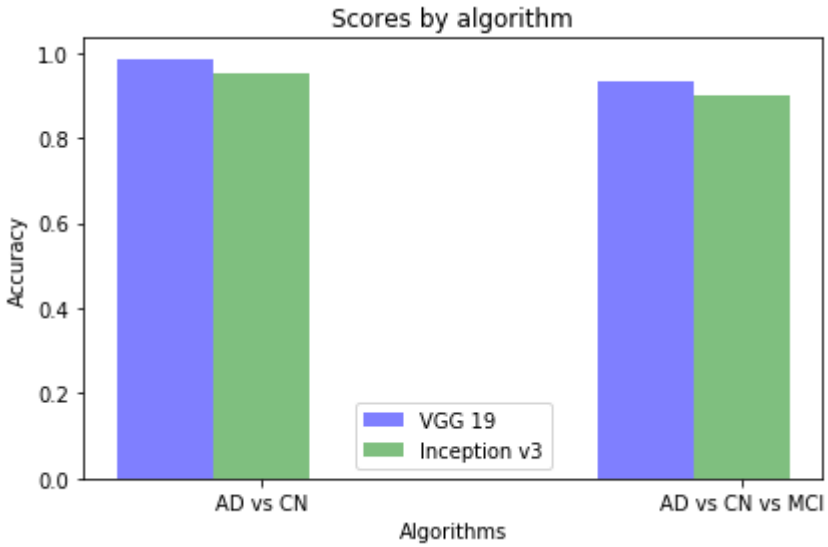

Figure 5.1: Scores by Algorithm Result in terms of VGG19 Vs Inception V3

We used ReLu activation function throughout multiple models. We ran AD vs CN through VGG19 and got 78% validation accuracy on 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 iteration. Next two set of iteration gave 86% and 95%. Then 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 iteration, after that next five iteration 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 next five iteration and 89% accuracy achieved for last set (figure 5.1). 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% (figure 5.2).

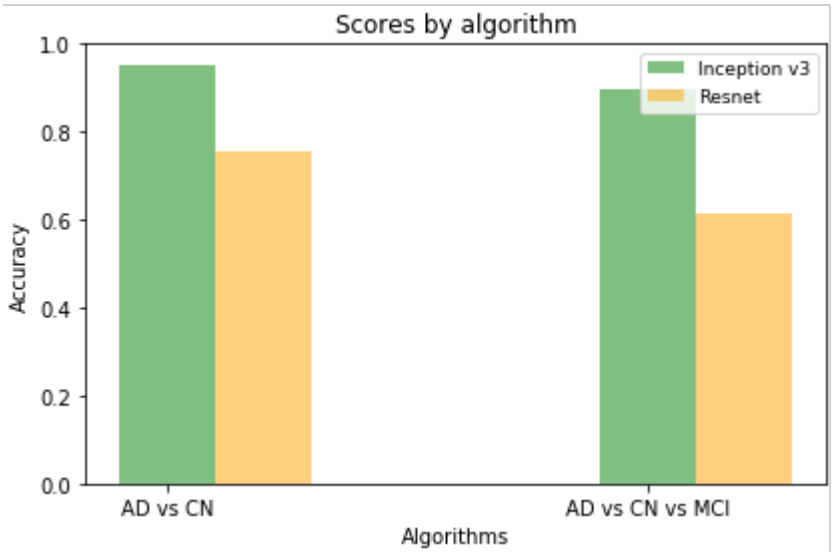

Figure 5.2: Scores by Algorithm Result in terms of ResNet Vs Inception V3

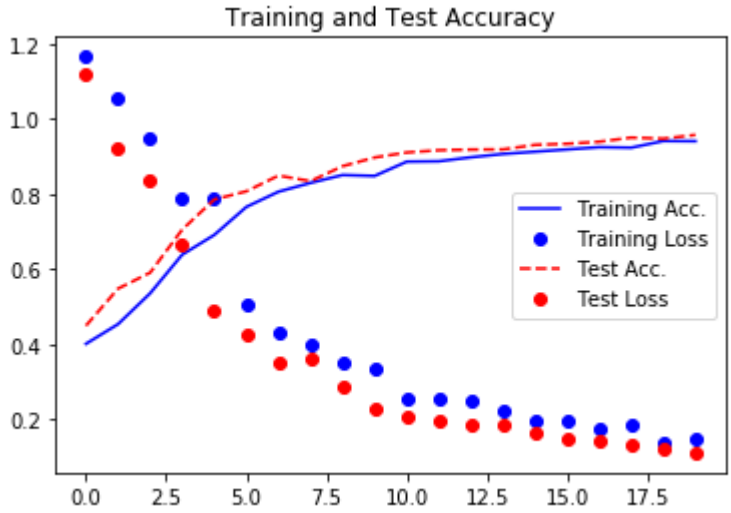

Figure 5.3: Train Vs Test Accuracy Result

5.0.2 Discussion

From our results we can state that we achieved an excellent result on our models. Our decision to drop out multi section layer from image set had an great impact on the transfer learning process. Though we had lower accuracy with Inception V3 model. Our train and test graph shows the learning curve (figure 5.3). The loss is decreasing throughout the model while model accuracy was getting higher. we ran 15 Epoch and stopped early cause the dataset was limited to 50 thousand. Therefore we could eliminate the chances of over fitting. If we had more data incorporated with our research domain, our precision level could increase exponentially. Our validation accuracy was higher than categorical accuracy which contrasts general machine learning concepts. As generally testing accuracy should be higher than test accuracy. however the results differs in our simulation because we used 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 during each iteration. From fig 5.2 we can see that the model improve significantly during first 10 epoch or so. However the improvement became less and less noticeable after the 10th epoch, so the training was stopped after 15th epoch (figure 5.4). As the improvement per epoch was almost ignoreable.
Our proposed model is different from existing methods because we have used transfer learning. This prevents the chances of getting stuck into local optima. It gives faster convergence. As transfer learning enables us to acquire decent accuracy right from the first epoch we can achieve good result with less hyperparameter tuning.

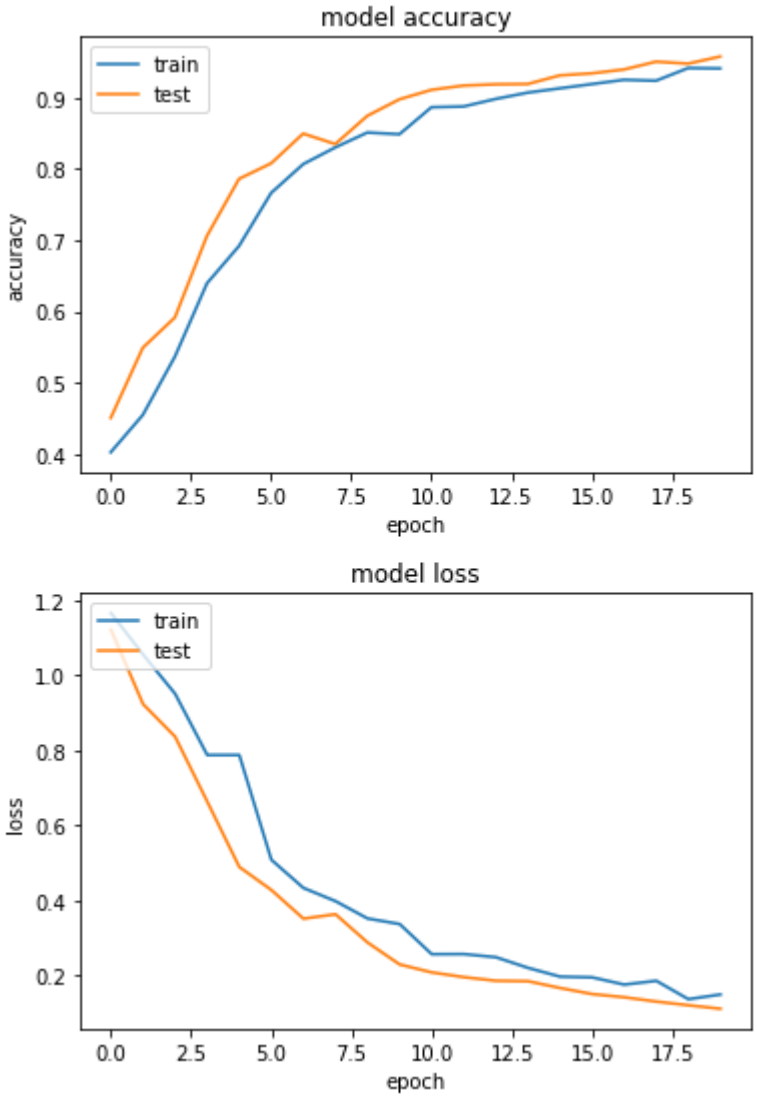

Figure 5.4: Model Accuracy Vs Model Loss

Chapter 6

Conclusion and Future Work

6.1 Conclusion

To conclude, from our experiment we can see that deep learning can work very efficiently for classification of Alzheimer 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 very crude way to solve a problem and give a dynamic results towards the research domain. 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.

6.2 Challenges

While working in this research, we have faced some challenges. One of them is we did not get to run more epoch as there was a lack of time. Another one is result can be biased on MCI because in the dataset, the number of MCI data was much more than the number of AD data.

6.3 Future Work

In our research we have collected 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 bio markers of Alzheimer disease for extracting the main features of the MRI images. However, when we have studied deeply, we have come to know that Alzheimer also has an impact on eye. The Alzheimer effected people's eye region can be differentiated from cognitively normal people's eye region. The genetic bio markers that are responsible for AD were also selected and categorized. For this process, we will be needing OCT image dataset of AD effected people which is not currently available. In near future, we are hoping that the dataset will be generated and we will continue our research on this area as eye bio markers are an important perspective in Alzheimer 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.

Bibliography

1. C. A. Taylor, S. F. Greenlund, L. C. McGuire, H. Lu, and J. B. Croft, from alzheimer's disease—united states, 1999-2014", MMWR. Morbidity and mortality weekly report, vol. 66, no. 20, p. 521, 2017.

2. A. Rahman, F. Salam, M. A. Islam, A. Anwarullah, M. R. Islam, M. N. A. Miah, U. K. Saha, and Z. Ali, s disease-an update", Bangladesh Journal of Neuroscience, vol. 28, no. 1, pp. 52-58, 2012.

3. J. Xu, K. D. Kochanek, S. L. Murphy, and B. Tejada-Vera, : Final data for 2014", 2016.

4. A. Association et al., "2017 alzheimer's disease facts and figures", Alzheimer's Dementia, vol. 13, no. 4, pp. 325-373, 2017.

5. A. Rahman, F. Salam, M. Islam, A. Anwarullah, M. Islam, M. Miah, U. Saha, and Z. Ali, disease - an update", Bangladesh Journal of Neuro- science, vol. 28, no. 1, pp. 52-58, Dec. 2013. doi: 10.3329/bjn.v28i1.17193. [Online]. Available: https ://www.banglajol . info/index.php/BJN/article/ view/17193.

6. M. Rahman, A. Tajmim, M. Ali, M. Sharif, et al., and current status of alzheimer's disease in bangladesh", Journal of Alzheimer's disease reports, vol. 1, no. 1, pp. 27-42, 2017.

7. V. L. Villemagne, S. Burnham, P. Bourgeat, B. Brown, K. A. Ellis, O. Salvado, C. Szoeke, S. L. Macaulay, R. Martins, P. Maru, et al.,

   deposition, neurodegeneration, and cognitive decline in sporadic alzheimer's disease: A prospective cohort study", The Lancet Neurology, vol. 12, no. 4, pp. 357-367, 2013.

8. A. J. Saykin, L. Shen, T. M. Foroud, S. G. Potkin, S. Swaminathan, S. Kim, S. L. Risacher, K. Nho, M. J. Huentelman, D. W. Craig, et al., 's disease neuroimaging initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans", Alzheimer's dementia, vol. 6, no. 3, pp. 265-273, 2010.

9. L. Bertram, M. B. McQueen, K. Mullin, D. Blacker, and R. E. Tanzi, meta-analyses of alzheimer disease genetic association studies: The alzgene database", Nature genetics, vol. 39, no. 1, p. 17, 2007.

10. A. Alzheimer's, 015 alzheimer's disease facts and figures.", Alzheimer's dementia: the journal of the Alzheimer's Association, vol. 11, no. 3, p. 332, 2015.

11. H. Hanyu, T. Sato, K. Hirao, H. Kanetaka, T. Iwamoto, and K. Koizumi, progression of cognitive deterioration and regional cerebral blood ow patterns in alzheimer's disease: A longitudinal spect study", Journal of the Neurological Sciences, vol. 290, no. 1, pp. 96-101, 2010, issn: 0022-510X. doi: https://doi.org/10.1016/j.jns.2009.10.022. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022510X0900940X.

12. K. R. Gray, R. Wolz, R. A. Heckemann, P. Aljabar, A. Hammers, and D. Rueckert, -region analysis of longitudinal fdg-pet for the classification of alzheimer's disease", NeuroImage, vol. 60, no. 1, pp. 221-229, 2012, issn: 1053- 8119. doi: https://doi.org/10.1016/j.neuroimage.2011.12.071. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S105381191101473X.

13. N. Schuff, N. Woerner, L. Boreta, T. Kornfield, L. M. Shaw, J. Q. Trojanowski, P. M. Thompson, J. Jack C. R., M. W. Weiner, and the Alzheimer's; Disease Neuroimaging Initiative, of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers", Brain, vol. 132, no. 4, pp. 1067-1077, Feb. 2009, issn: 0006-8950. doi: 10 . 1093 / brain/awp007. eprint: http://oup.prod.sis.lan/brain/article-pdf/132/4/1067/ 16694927/awp007.pdf. [Online]. Available: https://doi.org/10.1093/brain/ awp007.

14. N. Schu, N. Woerner, L. Boreta, T. Korn

    eld, L. M. Shaw, J. Q. Tro janowski, P. M. Thompson, J. Jack C. R., M. W. Weiner, and the Alzheimer's; Disease Neuroimaging Initiative, of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers", Brain, vol. 132, no. 4, pp. 1067-1077, Feb. 2009, issn: 0006-8950. doi: 10 . 1093 / brain/awp007. eprint: http://oup.prod.sis.lan/brain/article-pdf/132/4/1067/ 16694927/awp007.pdf. [Online]. Available: https://doi.org/10.1093/brain/ awp007.

15. J. Gorriz, F. Segovia, J. Ramirez, A. Lassl, and D. Salas-Gonzalez, based spect image classification for the diagnosis of alzheimer's disease", Ap- plied Soft Computing, vol. 11, no. 2, pp. 2313-2325, 2011, The Impact of Soft Computing for the Progress of Artificial Intelligence, issn: 1568-4946. doi: https : / / doi . org / 10 . 1016 / j . asoc . 2010 . 08 . 012. [Online]. Available: http : //www.sciencedirect.com/science/article/pii/S1568494610002140.

16. K. R. Gray, R. Wolz, R. A. Heckemann, P. Aljabar, A. Hammers, and D. Rueckert, -region analysis of longitudinal fdg-pet for the classification of alzheimer's disease", NeuroImage, vol. 60, no. 1, pp. 221-229, 2012, issn: 1053- 8119. doi: https://doi.org/10.1016/j.neuroimage.2011.12.071. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S105381191101473X.

17. A. S. Lundervold and A. Lundervold, overview of deep learning in medical imaging focusing on mri", Zeitschrift f Ar Medizinische Physik, vol. 29, no. 2, pp. 102-127, 2019, issn: 0939-3889. doi: https://doi.org/10.1016/j.zemedi.2018.11.002. [Online]. Available: http ://www. sciencedirect .com/science/article/pii/S0939388918301181.

18. I. Beheshti and H. Demirel, distribution function-based classifi- cation of structural mri for the detection of alzheimer's disease", Computers in Biology and

Medicine, vol. 64, pp. 208-216, 2015, issn: 0010-4825. doi: https ://doi .org/10.1016/j .compbiomed.2015.07.006. [Online]. Available: http://www.sciencedirect.com/s

19. G. Papakostas, A. Savio, M. Gra na, and V. Kaburlasos, lattice computing approach to alzheimer's disease computer assisted diagnosis based on mri data", Neurocomputing, vol. 150, pp. 37-42, 2015, Bioinspired and knowledge based techniques and applications The Vitality of Pattern Recognition and Image Analysis Data Stream Classification and Big Data Analytics, issn: 0925-2312. doi: https://doi.org/10.1016/j.neucom.2014.02.076. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231214012454.

20. Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, and T.-F. Yuan, of subjects and brain regions related to alzheimer's disease using 3d mri scans based on eigenbrain and machine learning", Frontiers in Com- putational Neuroscience, vol. 9, p. 66, 2015, issn: 1662-5188. doi: 10.3389/fncom.2015.00066. [Online]. Available: https://www.frontiersin.org/article/10.3389/fncom.2015.00066.

21. G. Papakostas, A. Savio, M. Gra na, and V. Kaburlasos, lattice computing approach to alzheimer's disease computer assisted diagnosis based on mri data", Neurocomputing, vol. 150, pp. 37-42, 2015, Bioinspired and knowledge based techniques and applications The Vitality of Pattern Recognition and Image Analysis Data Stream Classification and Big Data Analytics, issn: 0925-2312. doi: https://doi.org/10.1016/j.neucom.2014.02.076. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231214012454.

22. Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, and T.-F. Yuan, of subjects and brain regions related to alzheimer's disease using 3d mri scans based on eigenbrain and machine learning", Frontiers in Com- putational Neuroscience, vol. 9, p. 66, 2015, issn: 1662-5188. doi: 10.3389/fncom.2015.00066. [Online]. Available: https://www.frontiersin.org/article/10.3389/fncom.2015.00066.

23. A. H. Andersen, W. S. Rayens, Y. Liu, and C. D. Smith, least squares for discrimination in fmri data", Magnetic Resonance Imaging, vol. 30, no. 3, pp. 446-452, 2012, issn: 0730-725X. doi: https://doi.org/10.1016/j.mri.2011.11.001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0730725X11004565.

24. Y. Fan, S. M. Resnick, X. Wu, and C. Davatzikos, and functional biomarkers of prodromal alzheimer's disease: A high-dimensional pattern classification study", NeuroImage, vol. 41, no. 2, pp. 277-285, 2008, issn: 1053- 8119. doi: https://doi.org/10.1016/j.neuroimage.2008.02.043. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S105381190800178X.

25. M. Gra na, M. Termenon, A. Savio, A. Gonzalez-Pinto, J. Echeveste, J. Perez, and A. Besga, aided diagnosis system for alzheimer disease using brain diffusion tensor imaging features selected by pearson's correlation", Neuroscience Letters, vol. 502, no. 3, pp. 225-229, 2011, issn: 0304-3940. doi: https ://doi . org/10.1016/j . neulet .2011.07.049. [Online]. Available: http : //www.sciencedirect.com/science/article/pii/S0304394011011268.

26. D. B. Henley, S. A. Dowsett, Y.-F. Chen, H. Liu-Seifert, J. D. Grill, R. S. Doody, P. Aisen, R. Raman, D. S. Miller, A. M. Hake, et al., 's disease progression by geographical region in a clinical trial setting", Alzheimer's research therapy, vol. 7, no. 1, p. 43, 2015.

27. J. Islam and Y. Zhang, ensemble of deep convolutional neural networks for alzheimer's disease detection and classification", arXiv preprint arXiv:1712.01675, 2017.

28. N. C. Fox, E. K. Warrington, P. A. Freeborough, P. Hartikainen, A. M. Kennedy, J. M. Stevens, and M. N. Rossor, hippocampal atrophy in Alzheimer's disease: A longitudinal MRI study", Brain, vol. 119, no. 6, pp. 2001-2007, Dec. 1996, issn: 0006-8950. doi: 10.1093/brain/119.6.2001. eprint: http://oup.prod.sis.lan/brain/article-pdf/119/6/2001/1265727/119-6- 2001.pdf. [Online]. Available: https://doi.org/10.1093/brain/119.6.2001.

29. C. R. Jack, R. C. Petersen, Y. C. Xu, S. C. Waring, P. C. O'Brien, E. G. Tangalos, G. E. Smith, R. J. Ivnik, and E. Kokmen, temporal atrophy on mri in normal aging and very mild alzheimer's disease", Neurology, vol. 49, no. 3, pp. 786-794, 1997, issn: 0028-3878. doi: 10.1212/WNL.49.3.786. eprint: https://n.neurology.org/content/49/3/786.full.pdf. [Online]. Available: https: //n.neurology.org/content/49/3/786.

30. A. S. Lundervold and A. Lundervold, overview of deep learning in medical imaging focusing on mri", Zeitschrift f Ar Medizinische Physik, vol. 29, no. 2, pp. 102-127, 2018, issn: 0939-3889. doi: https://doi.org/10.1016/j.zemedi.2018.11.002.

31. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks", in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2014, pp. 1725-1732.

32. S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition", IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 1, pp. 221-231, 2012.

33. A. S. Lundervold and A. Lundervold, overview of deep learning in medical imaging focusing on mri", Zeitschrift fur Medizinische Physik, vol. 29, no. 2, pp. 102-127, 2019.

34. H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning", IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285-1298, May 2016, issn: 0278-0062. doi: 10.1109/TMI.2016.2528162.

35. L. O. Chua and T. Roska, cnn paradigm", IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, vol. 40, no. 3, pp. 147-156, Mar. 1993, issn: 1057-7122. doi: 10.1109/81.222795.

36. A. S. Lundervold and A. Lundervold, overview of deep learning in medical imaging focusing on mri", Zeitschrift fur Medizinische Physik, vol. 29, no. 2, pp. 102-127, 2017.

37. S. Sonoda and N. Murata, network with unbounded activation functions is universal approximator", Applied and Computational Harmonic Anal- ysis, vol. 43, no. 2, pp. 233-268, 2017.

38. M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, feedforward networks with a nonpolynomial activation function can approximate any function", Neural networks, vol. 6, no. 6, pp. 861-867, 1993.

39. Sonoda and N. Murata, network with unbounded activation functions is universal approximator", Applied and Computational Harmonic Anal- ysis, vol. 43, no. 2, pp. 233-268, 2017.

40. A. S. Lundervold and A. Lundervold, overview of deep learning in medical imaging focusing on mri", Zeitschrift fur Medizinische Physik, vol. 29, no. 2, pp. 102-127, 2019.

41. J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, for simplicity: The all convolutional net", arXiv preprint arXiv:1412.6806, 2014.

42. S. Wager, S. Wang, and P. S. Liang, training as adaptive regularization", in Advances in Neural Information Processing Systems 26, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2013, pp. 351-359. [Online]. Available: http://papers.nips.cc/paper/4882-dropout-training-as-adaptive-regularization.pdf.

43. J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, for simplicity: The all convolutional net", arXiv preprint arXiv:1412.6806, 2014.

44. S. Ioffe and C. Szegedy, normalization: Accelerating deep network training by reducing internal covariate shift", arXiv preprint arXiv:1502.03167, 2015.

45. W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, for transfer learning", in Proceedings of the 24th international conference on Machine learning, ACM, 2007, pp. 193-200.

46. R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, -taught learning: Transfer learning from unlabeled data", in Proceedings of the 24th interna- tional conference on Machine learning, ACM, 2007, pp. 759-766.

47. S. J. Pan, J. T. Kwok, Q. Yang, et al., learning via dimensionality reduction.", in AAAI, vol. 8, 2008, pp. 677-682.

48. P. Y. Simard, D. Steinkraus, J. C. Platt, et al., practices for convolutional neural networks applied to visual document analysis.", in Icdar, vol. 3, 2003.

49. R. J. Williams and D. Zipser, learning algorithm for continually running fully recurrent neural networks", Neural computation, vol. 1, no. 2, pp. 270-280, 1989.

50. F. Gruau et al., network synthesis using cellular encoding and the genetic algorithm.", 1994.

51. K. Simonyan and A. Zisserman, deep convolutional networks for largescale image recognition", arXiv preprint arXiv:1409.1556, 2014.

52. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, large scale visual recognition challenge", International Journal of Computer Vision, vol. 115, no. 3, pp. 211-252, Dec. 2015, issn: 1573-1405. doi: 10.1007/s11263-015-0816-y. [Online]. Available: https://doi.org/10.1007/s11263-015-0816-y.

53. H. Ergun and M. Sert, deep convolutional networks for large scale visual concept classification", in 2016 IEEE Second International Conference on Multimedia Big Data (BigMM), IEEE, 2016, pp. 210-213.

54. M. Symms, H. R. Jager, K. Schmierer, and T. A. Yousry, review of structural magnetic resonance neuroimaging", Journal of Neurology, Neurosurgery Psychiatry, vol. 75, no. 9, pp. 1235-1244, 2004, issn: 0022-3050. doi: 10.1136/jnnp.2003.032714. eprint: https://jnnp.bmj.com/content/75/9/1235. full.pdf. [Online]. Available: https://jnnp.bmj.com/content/75/9/1235.

55. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

56. He, K., Zhang, X., Ren, S., Sun, J. (2016, October). Identity mappings in deep residual networks. In European conference on computer vision (pp. 630-645). Springer, Cham.