8-13-2019

TOWARDS AI-ASSISTED DISEASE DIAGNOSIS: LEARNING DEEP FEATURE REPRESENTATIONS FOR MEDICAL IMAGE ANALYSIS

Jyoti Islam

Follow this and additional works at: https://scholarworks.gsu.edu/cs_diss

Recommended Citation
https://scholarworks.gsu.edu/cs_diss/152
TOWARDS AI-ASSISTED DISEASE DIAGNOSIS: LEARNING DEEP FEATURE REPRESENTATIONS FOR MEDICAL IMAGE ANALYSIS

by

JYOTI ISLAM

Under the Direction of Yanqing Zhang, PhD

ABSTRACT

Artificial Intelligence (AI) has impacted our lives in many meaningful ways. For our research, we focus on improving disease diagnosis systems by analyzing medical images using AI, specifically deep learning technologies. The recent advances in deep learning technologies are leading to enhanced performance for medical image analysis and computer-aided disease diagnosis. In this dissertation, we explore a major research area in medical image analysis - Image classification. Image classification is the process to assign an image a label from a fixed
set of categories. For our research, we focus on the problem of Alzheimer’s Disease (AD) diagnosis from 3D structural Magnetic Resonance Imaging (sMRI) and Positron Emission Tomography (PET) brain scans.

Alzheimer’s Disease is a severe neurological disorder. In this dissertation, we address challenges related to Alzheimer’s Disease diagnosis and propose several models for improved diagnosis. We focus on analyzing the 3D Structural MRI (sMRI) and Positron Emission Tomography (PET) brain scans to identify the current stage of Alzheimer’s Disease: Normal Controls (CN), Mild Cognitive Impairment (MCI), and Alzheimer’s Disease (AD). This dissertation demonstrates ways to improve the performance of a Convolutional Neural Network (CNN) for Alzheimer’s Disease diagnosis. Besides, we present approaches to solve the class-imbalance problem and improving classification performance with limited training data for medical image analysis. To understand the decision of the CNN, we present methods to visualize the behavior of a CNN model for disease diagnosis. As a case study, we analyzed brain PET scans of AD and CN patients to see how CNN discriminates among data samples of different classes.

Additionally, this dissertation proposes a novel approach to generate synthetic medical images using Generative Adversarial Networks (GANs). Working with the limited dataset and small amount of annotated samples makes it difficult to develop a robust automated disease diagnosis model. Our proposed model can solve such issue and generate brain MRI and PET images for three different stages of Alzheimer’s Disease - Normal Control (CN), Mild Cognitive Impairment (MCI), and Alzheimer’s Disease (AD). Our proposed approach can be generalized to create synthetic data for other medical image analysis problems and help to develop better disease diagnosis model.
INDEX WORDS: Medical Imaging, Deep Learning, Alzheimer’s Disease, MRI, PET, Brain Imaging, Synthetic Data Generation, Visualization, Convolutional Neural Network, Generative Adversarial Networks, Computer-aided-diagnosis
TOWARDS AI-ASSISTED DISEASE DIAGNOSIS: LEARNING DEEP FEATURE REPRESENTATIONS FOR MEDICAL IMAGE ANALYSIS

by

JYOTI ISLAM

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2019
DEDICATION

I dedicate this work to my family, especially my always supporting husband (Md Modasshir), mom (Gulnaher Begum), and sister (Dyuti Islam) for their love, encouragement, and endless support. This dissertation was only possible because of their presence in my life.
ACKNOWLEDGEMENTS

This dissertation is the result of research carried out over many years. During this time, I have received support, encouragement and smiles from many great people. I would have never been able to finish my dissertation without them, and here, I would like to express my gratitude.

I want to express my sincere gratitude to my advisor Dr. Yanqing Zhang for his support and invaluable guidance throughout my study. His knowledge, perceptiveness, and innovative ideas have guided me throughout my graduate study. It was a great privilege and honor to work and study under his guidance.

I also present my words of gratitude to the other members of my dissertation committee, Dr. Saeid Belkasim, Dr. Jonathan Shihao Ji, and Dr. Walter Wilczynski for their advice and valuable time spent in reviewing the material. Their guidance helped me in all the time of research and writing of this dissertation.

I want to thank other professors of computer science department who offered me advice and moral support throughout my graduate studies. I express my gratitude towards the staff members in our department for their cordial help during my PhD life.

I am thankful to the professors of the computer science department from the University of Dhaka, Bangladesh, who shaped my life during my undergraduate years. I express my gratitude to my mentors, and colleagues from Together Initiatives Limited, Samsung R & D Institute Bangladesh and Blackline Systems, Inc., for their whole-hearted support in everything.

I am forever grateful to my husband, Md Modasshir, for his support, encouragement, patience and unwavering love during my PhD. He is undeniably the bedrock of my achievements in the past years. He believed in me even when I didn’t believe in myself and pushed me to go forward.

I am truly indebted to my family back home for their support throughout my study.
Without their efforts, sacrifice, and struggle, I would not be who I am today. I express my gratitude to my father Md Anwar Hossain, my mother Gulnaher Begum, my sisters Dipti Islam, Dyuti Islam and Ahona Islam for their love and support during my studies.

I express my sincere gratitude to my Atlanta family who supported and loved me unconditionally throughout my PhD journey. I would also like to thank all my friends, extended family members, and colleagues for their love and support.

Last but not least, it is a pleasure to thank everybody who made this dissertation possible, as well as express my apologies that I could not mention everyone individually.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ......................................................... v

LIST OF TABLES ............................................................... xi

LIST OF FIGURES ............................................................. xii

LIST OF ABBREVIATIONS ..................................................... xvi

Chapter 1 INTRODUCTION ..................................................... 1
  1.1 AI-Assisted Disease Diagnosis ........................................ 1
  1.2 Medical Image Analysis ................................................ 1
  1.3 Classification in Medical Imaging ...................................... 2
     1.3.1 Alzheimer’s Disease Diagnosis .................................. 2
  1.4 Feature Learning ........................................................ 4
  1.5 Deep Learning .......................................................... 5
  1.6 Objective ............................................................... 5
  1.7 Motivation ............................................................... 5
  1.8 Challenges ............................................................. 6
  1.9 Contribution ........................................................... 7
  1.10 List of Publications ................................................... 8
  1.11 Thesis Outline ........................................................ 10

Chapter 2 BACKGROUND STUDY ............................................. 11
  2.1 Learning Algorithms ................................................... 11
  2.2 Neural Network ........................................................ 13
  2.3 Convolutional Neural Network ........................................ 13
     2.3.1 Local Connectivity ................................................ 15
2.3.2 Shared Weights ........................................... 16
2.3.3 Pooling .................................................... 16
2.4 Generative Adversarial Networks ......................... 17
2.5 Magnetic Resonance Imaging ............................... 19
2.5.1 Structural MRI (sMRI) .................................. 20
2.6 Positron Emission Tomography (PET) .................... 21

Chapter 3 REVIEW OF THE STATE-OF-THE-ART ............... 24
3.1 Alzheimer’s Disease Diagnosis ............................. 24
3.2 Medical Image Synthesis ................................... 28
3.3 CNN Visualization .......................................... 29

Chapter 4 A NOVEL DEEP LEARNING BASED MULTI-CLASS
CLASSIFICATION METHOD FOR ALZHEIMER’S DISEASE DIAGNOSIS .......... 30
4.1 Introduction ................................................. 30
4.2 Method ..................................................... 30
4.3 Data ......................................................... 32
4.4 Experiments ............................................... 33
   4.4.1 Results ............................................... 34
4.5 Summary ................................................... 34

Chapter 5 ALZHEIMER’S DISEASE DIAGNOSIS USING AN ENSEMBLE SYSTEM OF DEEP CONVOLUTIONAL NEURAL NETWORKS ............... 36
5.1 Introduction ............................................... 36
5.2 Method ..................................................... 37
   5.2.1 Data Selection ......................................... 37
   5.2.2 Data Augmentation .................................... 38
5.2.3 Network Architecture .............................................. 38
5.3 Experiments .......................................................... 46
  5.3.1 Results .............................................................. 48
5.4 Summary .............................................................. 51

Chapter 6  DEEP CONVOLUTIONAL NEURAL NETWORKS FOR AUTOMATED DIAGNOSIS OF ALZHEIMER’S DISEASE AND MILD COGNITIVE IMPAIRMENT ............. 52
  6.1 Introduction .......................................................... 52
  6.2 Method ............................................................... 53
    6.2.1 Data Selection .................................................. 53
    6.2.2 Data Preprocessing ........................................... 54
    6.2.3 Data Augmentation ........................................... 55
    6.2.4 Network Architecture ....................................... 56
  6.3 Experiments .......................................................... 57
    6.3.1 Results ............................................................ 58
  6.4 Summary ............................................................. 59

Chapter 7  GAN-BASED SYNTHETIC BRAIN PET IMAGE GENERATION .................................................. 60
  7.1 Introduction .......................................................... 60
  7.2 Method ............................................................... 63
    7.2.1 Data Selection .................................................. 63
    7.2.2 Generative Adversarial Networks ............................ 63
    7.2.3 Deep Convolutional Generative Adversarial Networks (DC-GANs) ............................................. 66
    7.2.4 Proposed Model ................................................ 67
  7.3 Experiments and Results .......................................... 70
  7.4 Summary ............................................................. 72
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 4.1</td>
<td>Confusion Matrix</td>
<td>34</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Five-fold Cross Validation Performance Accuracy comparison</td>
<td>34</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Classification Performance of M&lt;sub&gt;1&lt;/sub&gt; Model</td>
<td>47</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Classification Performance of M&lt;sub&gt;2&lt;/sub&gt; Model</td>
<td>48</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Classification Performance of M&lt;sub&gt;3&lt;/sub&gt; Model</td>
<td>48</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>Classification Performance of M&lt;sub&gt;4&lt;/sub&gt; Model</td>
<td>48</td>
</tr>
<tr>
<td>Table 5.5</td>
<td>Classification Performance of M&lt;sub&gt;5&lt;/sub&gt; Model</td>
<td>49</td>
</tr>
<tr>
<td>Table 5.6</td>
<td>Performance of the Proposed Ensembled Model</td>
<td>49</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>Impact of Different Factors on Proposed CNN</td>
<td>57</td>
</tr>
<tr>
<td>Table 6.2</td>
<td>Impact of Different CNN Architecture on Proposed Diagnosis Framework</td>
<td>57</td>
</tr>
<tr>
<td>Table 6.3</td>
<td>Comparison with the State-of-the-Art. ‘–’ indicates that result was not reported by the authors</td>
<td>58</td>
</tr>
<tr>
<td>Table 8.1</td>
<td>Brain Area Relevance for CN, MCI, and AD Patient for Each Visualization Method</td>
<td>88</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

Table of Contents:

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Sample Brain MRI Images from ADNI Database Presenting Different Stages of Alzheimer’s Disease. (a) Normal/healthy controls (CN); (b) Mild Cognitive Impairment (MCI); (c) Alzheimer’s Disease (AD).</td>
<td>3</td>
</tr>
<tr>
<td>2.1</td>
<td>Relation Among AI, Machine Learning, and Deep Learning.</td>
<td>12</td>
</tr>
<tr>
<td>2.2</td>
<td>Block Diagram of a Neural Network.</td>
<td>12</td>
</tr>
<tr>
<td>2.3</td>
<td>Illustration of a Convolutional Neural Network.</td>
<td>14</td>
</tr>
<tr>
<td>2.4</td>
<td>Local Connectivity of CNN.</td>
<td>15</td>
</tr>
<tr>
<td>2.5</td>
<td>Feature Map.</td>
<td>16</td>
</tr>
<tr>
<td>2.6</td>
<td>Max-Pooling Operation. Each Max is Taken Over 4 Numbers Arranged in 2x2 Square.</td>
<td>17</td>
</tr>
<tr>
<td>2.7</td>
<td>The Major Components of A Magnetic Resonance Imaging System [1].</td>
<td>18</td>
</tr>
<tr>
<td>2.8</td>
<td>T1 and T2 Weighted Images of Brain [2].</td>
<td>19</td>
</tr>
<tr>
<td>2.9</td>
<td>Structural MRI Images Presenting (a) Normal Control; (b) MCI; (c) AD.</td>
<td>20</td>
</tr>
<tr>
<td>2.10</td>
<td>Positron Emission Tomography (PET) Scanning Procedure [3].</td>
<td>21</td>
</tr>
<tr>
<td>2.11</td>
<td>Example of Brain PET Images (a) Normal Control (b) MCI (c) AD.</td>
<td>22</td>
</tr>
<tr>
<td>4.1</td>
<td>Example of Different Brain MRI Images Presenting Different AD Stage. (a) Nondemented; (b) very mild dementia; (c) mild dementia; (d) moderate dementia.</td>
<td>30</td>
</tr>
<tr>
<td>4.2</td>
<td>Block Diagram of Proposed Alzheimer’s Disease Detection and Classification Framework.</td>
<td>31</td>
</tr>
<tr>
<td>4.3</td>
<td>Sample Images From OASIS Dataset.</td>
<td>33</td>
</tr>
<tr>
<td>5.1</td>
<td>Illustration of Dense Connectivity with a 5-layer Dense Block.</td>
<td>39</td>
</tr>
<tr>
<td>5.2</td>
<td>Illustration of Two Dense Blocks and Their Intermediate Connection.</td>
<td>39</td>
</tr>
</tbody>
</table>
Figure 5.3 Block Diagram of Proposed Alzheimer’s Disease Diagnosis Framework. ................................................................. 40
Figure 5.4 Block Diagram of Individual Model M_4. ................................................................. 45
Figure 5.5 Block Diagram of Individual Model M_5. ................................................................. 46
Figure 5.6 Performance Comparison of the Proposed Model and the Variants. ................. 49
Figure 5.7 Performance Comparison of the Proposed Model and the Baseline Deep CNNs. ................................................................. 50
Figure 5.8 Comparison of Accuracy on the OASIS Dataset. ................................................................. 50
Figure 6.1 3D Brain MRI Preprocessing module. ................................................................. 54
Figure 6.2 Skull-Stripped MRI Slices Presenting Different AD Stages. (a)-(c) CN; (d)-(f) MCI; (g)-(i) AD. ................................................................. 55
Figure 6.3 Block Diagram of the Proposed Alzheimer’s Disease Diagnosis Framework. ................................................................. 55
Figure 6.4 Deep CNN Architecture Used for the Proposed Alzheimer’s Disease Diagnosis Framework. ................................................................. 56
Figure 7.1 Example of Brain PET images (a) Sagittal View (b) Coronal View (c) Axial View. ................................................................. 61
Figure 7.2 Proposed Synthetic Brain PET Image Generator. ................................................................. 62
Figure 7.3 Generator Architecture of the Proposed Model. ................................................................. 63
Figure 7.4 Discriminator Architecture of the Proposed Model. ................................................................. 64
Figure 7.5 Visualization of the Generator Output in the Training Process. ................................................................. 65
Figure 7.6 Real and Synthetic Brain PET Images of Normal Patient (a) Real (b) Synthetic. ................................................................. 66
Figure 7.7 Real and Synthetic Brain PET Images of MCI Patient (a) Real (b) Synthetic. ................................................................. 67
Figure 7.8 Real and Synthetic Brain PET Images of AD Patient (a) Real (b) Synthetic. ................................................................. 69
Figure 7.9  2D-Histograms of the Synthetic and Real Images. Top Row: 2D-Histogram of Real images. Bottom Row: 2D-Histogram of Synthetic Images.  

Figure 8.1  Example of Brain PET Images (a) Normal Control (b) Mild Cognitive Impairment (c) Alzheimer’s Disease.  

Figure 8.2  Skull-Stripped PET Slices.  

Figure 8.3  Alzheimer’s Disease diagnosis framework using PET data.  

Figure 8.4  3D CNN Classifier for Alzheimer’s Disease and Mild Cognitive Impairment diagnosis using PET data.  

Figure 8.5  Relevance Heatmaps for Sensitivity Analysis (Backpropagation) Method Averaged Over CN Patients.  

Figure 8.6  Relevance Heatmaps for Sensitivity Analysis (Backpropagation) Method Averaged Over MCI Patients.  

Figure 8.7  Relevance Heatmaps for Sensitivity Analysis (Backpropagation) Method Averaged Over AD Patients.  

Figure 8.8  Relevance Heatmaps for Guided Backpropagation Method Averaged Over CN Patients.  

Figure 8.9  Relevance Heatmaps for Guided Backpropagation Method Averaged Over MCI Patients.  

Figure 8.10  Relevance Heatmaps for Guided Backpropagation Method Averaged Over AD Patients.  

Figure 8.11  Relevance Heatmaps for Occlusion Method Averaged Over CN Patients.  

Figure 8.12  Relevance Heatmaps for Occlusion Method Averaged Over MCI Patients.  

Figure 8.13  Relevance Heatmaps for Occlusion Method Averaged Over AD Patients.
Figure 8.14 Relevance Heatmaps for Brain Area Occlusion Method Averaged Over CN Patients. ...................................................... 89
Figure 8.15 Relevance Heatmaps for Brain Area Occlusion Method Averaged Over MCI Patients. ..................................................... 90
Figure 8.16 Relevance Heatmaps for Brain Area Occlusion Method Averaged Over AD Patients. ....................................................... 91
Figure 8.17 Relevance Heatmaps for Layer-wise Relevance Propagation (LRP) Method Averaged Over CN Patients. .......................... 92
Figure 8.18 Relevance Heatmaps for Layer-wise Relevance Propagation (LRP) Method Averaged Over MCI Patients. ....................... 93
Figure 8.19 Relevance Heatmaps for Layer-wise Relevance Propagation (LRP) Method Averaged Over AD Patients. ....................... 94
Figure 8.20 Visualization Comparison of the Relevance Heatmaps. (a) Sensitivity Analysis (Backpropagation). (b) Guided Backpropagation. (c) Occlusion. (d) Brain Area Occlusion. (e) Layer-wise Relevance Propagation (LRP). ......................................................... 95
LIST OF ABBREVIATIONS

• GSU - Georgia State University
• CS - Computer Science
• AI - Artificial Intelligence
• ML - Machine Learning
• DL - Deep Learning
• NN - Neural Network
• CNN - Convolutional Neural Network
• MRI - Magnetic Resonance Imaging
• sMRI - Structural Magnetic Resonance Imaging
• AD - Alzheimer’s Disease
• MCI - Mild Cognitive Impairment
• CN - Normal Control
• CAD - Computer Aided Diagnosis
• ReLU - Rectified Linear Unit
• 3D - Three Dimensional
• RF - Radio–frequency
• TR - Repetition Time
• TE - Time to Echo
• GM - Gray Matter
• WM - White Matter
• CSF - Cerebrospinal Fluid
• fMRI - Functional Magnetic Resonance Imaging
• PET - Position Emission Tomography
• SPECT - Single Photon Emission Computed Tomography
• DTI - Diffusion Tensor Imaging
• CT - Computed Tomography
• ROI - Region of Interest
• SVM - Support Vector Machine
• GLCM - Gray-Level Co-occurrence Matrix
• HC - Healthy Control
• PCA - Principal Component Analysis
• GAN - Generative Adversarial Network
• LRP - Layer-wise Relevance Propagation
• ADNI - Alzheimer’s Disease Neuroimaging Initiative
Chapter 1

INTRODUCTION

In this chapter, we present the general introduction to this thesis intended to allow a quick appraisal of its contents, contributions, supporting publications and structure.

1.1 AI-Assisted Disease Diagnosis

Artificial Intelligence (AI) has made substantial progress in recent days. AI is the area of computer science that aims to mimic human cognitive functions and emphasize the creation of intelligent machines. Today AI assisted models are performing incredible tasks with high accuracy at a massive scale. AI has enabled us to build tools that can learn from experience, adjust to new inputs and complete tasks with expert human-level performance. In this thesis, we aim to use AI for solving various problems in the field of medical image analysis focusing on improving performance for disease diagnosis.

Computer-aided diagnosis (CAD) refers to systems that provide information about disease assessment by interpreting medical images. AI can improve the performance of CAD systems with discoveries and judgments leading to faster and more accurate diagnosis. AI can handle a vast amount of information faster than any human with more accurate analysis. Besides, AI can improve the accuracy by learning itself and acquiring expertise comparable to specialists.

1.2 Medical Image Analysis

Medical image analysis focuses on extracting insights from images of biological tissue with computational analysis. Medical Imaging creates visual representations of body interior for clinical assessment and uses several imaging techniques such as Computer Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound, Radiography(X-Ray), Positron
emission tomography (PET) etc. The scope of medical image analysis ranges from clinical studies of medical imaging and disease in patients to neuroscience studies that target to find scientific information such as human brain structure and functionality. Several computational methods such as signal processing, machine learning, biophysics etc. are used to built applications for medical image analysis. In recent days deep learning techniques are helping to identify, classify and qualify patterns in medical images with the advantages of learning hierarchical feature representations directly from data instead of using hand-crafted features.

1.3 Classification in Medical Imaging

Image classification aims to assign an image to categories or classes of interest by analyzing its contents. In computer-aided diagnosis, image classification plays a significant role. Classification in medical imaging refers to the task of analyzing an input image data and assign it an output class indicating the presence of a disease or not. For example, classifying patients as having Alzheimer’s Disease based on 3D brain MRI data.

1.3.1 Alzheimer’s Disease Diagnosis

Alzheimer’s Disease (AD) is the most prevailing type of dementia. The prevalence of AD is estimated to be around 5% after 65 years old and is staggering 30% for more than 85 years old in developed countries. It is estimated that by 2050, around 0.64 Billion people will be diagnosed with AD [4]. Alzheimer’s Disease destroys brain cells causing people to lose their memory, mental functions and ability to continue daily activities. Initially, Alzheimer’s Disease affects the part of the brain that controls language and memory. As a result, AD patients suffer from memory loss, confusion, and difficulty in speaking, reading or writing. They often forget about their life and may not recognize their family members. They struggle to perform daily activities such as brushing hair or combing tooth. All these make AD patients anxious or aggressive or to wander away from home. Ultimately, AD destroys the part of the brain controlling breathing and heart functionality which lead to death.
There are three major stages in Alzheimer’s Disease - very mild, mild and moderate. Detection of Alzheimer’s Disease (AD) is still not accurate until a patient reaches moderate AD stage. For proper medical assessment of AD, several things are needed such as physical and neurobiological exams, Mini-Mental State Examination (MMSE), and patient’s detailed history. As AD is incurable, earlier diagnosis of AD can help for proper treatment. For our research, we consider the automated diagnosis of Alzheimer’s Disease in 3D structural MRI brain scans and Positron Emission Tomography (PET) brain scans. We conduct experiments using OASIS dataset and Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset for classification of the Alzheimer’s Disease (AD), Mild Cognitive Impairment (MCI) and CN (normal/healthy controls) to evaluate the proposed model. Mild Cognitive Impairment (MCI) combines the very mild and mild stages of Alzheimer’s disease. Fig. 1.1 shows some brain MRI images with different AD stages. As the disease progresses, Abnormal pro-

Figure (1.1) Sample Brain MRI Images from ADNI Database Presenting Different Stages of Alzheimer’s Disease. (a) Normal/healthy controls (CN); (b) Mild Cognitive Impairment(MCI); (c) Alzheimer’s Disease (AD).
teins (amyloid-$\beta$ [A$\beta$] and hyperphosphorylated tau) are accumulated in the brain of an AD patient. This abnormal protein accumulation leads to progressive synaptic, neuronal and axonal damage. The changes in the brain due to AD have a stereotypical pattern of early medial temporal lobe (entorhinal cortex and hippocampus) involvement, followed by progressive neocortical damage [5]. Such changes occur years before the AD symptoms appear. It looks like the toxic effects of hyperphosphorylated tau and/or amyloid-$\beta$ [A$\beta$] gradually erodes the brain, and when a clinical threshold is surpassed, amnestic symptoms start to develop. Structural MRI (sMRI) and Positron Emission Tomography (PET) can be used for measuring these progressive changes in the brain due to the AD. Our research work focuses on analyzing sMRI and PET data using deep learning model for Alzheimer’s Disease diagnosis.

1.4 Feature Learning

Feature learning or representation learning refers to a set of methods that allows a model to automatically identify the representations needed for feature detection or classification from raw data. Feature learning aims to find an appropriate representation of data to perform a machine learning task, eliminates the need to manual feature engineering and allows a machine to learn the features themselves. Real world data - images and video are complex and highly variable. For intelligent machines, it is necessary that they can discover useful features or representations from the raw data themselves. Traditional hand-crafted feature generation is expensive regarding time, cost and requires expert knowledge and human labor. Moreover, they do not generalize well. Efficient feature learning techniques can solve these issues and automate and generalize the learning process. Feature learning algorithms have two categories: supervised and unsupervised feature learning. In Supervised feature learning, the algorithm learns the feature from a labelled training dataset. On the other hand, in unsupervised feature learning, features are learned from unlabeled input data.
1.5 Deep Learning

Deep Learning is an area of Artificial Intelligence, specifically a part of Machine Learning Algorithms. Recently, deep learning models have been famous for their ability to learn feature representations from the input data. Deep learning networks use a layered, hierarchical structure to learn increasingly abstract feature representations from the data. Deep learning architectures learn simple, low-level features from the data and build complex high-level features in a hierarchy fashion. Deep learning technologies have demonstrated revolutionary performance in several areas, e.g., visual object recognition, human action recognition, natural language processing, object tracking, image restoration, denoising, segmentation tasks, audio classification, brain-computer interaction, etc. Nowadays deep learning techniques are being successfully applied for medical image analysis.

1.6 Objective

We developed computer-aided diagnosis models using deep learning technologies. We explored a significant area in medical image analysis - classification. For classification task as a case study, we developed automated models for Alzheimer’s disease diagnosis using brain MRI and PET data. We worked for solving the class-imbalance problem in medical data and developing efficient diagnosis models with limited training data. We propose solution to solve limited dataset problem by generating synthetic medical images. We present approaches to understand the decision of convolutional neural network to discriminate among samples of different disease class.

1.7 Motivation

AI can offer several advantages over traditional analytics, and clinical decision-making approaches. The health-care industry is a high priority sector where people expect the highest level of service. Proper diagnosis of disease is crucial to treatment and saving a patient’s life. For example, Alzheimer’s disease is not curable. But if we can diagnosis it at
an early stage, physicians can at least try to delay the harmful impact of the disease. On the other hand, diagnosis is expensive and requires specialized expertise. In many countries of the world, primary medical care is limited, and expert physicians are not available in time of need. If we can build computer-aided-diagnosis systems that can automatically analyze medical images and diagnosis disease with human-level expertise, it will reduce the cost of treatment dramatically. Early diagnosis of disease will save the lives of millions of people and bring exciting break-throughs in patient care.

While several research works have been done for analyzing natural images such as object classification, recognition, tracking, segmentation etc., research in medical image analysis is still limited. To mitigate the gap, we choose to focus on medical image analysis for developing better disease diagnosis systems.

1.8 Challenges

For our research, we plan to use deep learning technologies for medical image analysis. There are several challenges related to working with medical data. First, medical images are generally private and having access to those data is challenging and often impossible. Second, for natural image analysis such as object classification, millions of data are available to train an automated classifier. On the other hand, in disease classification, typically few hundreds data sample are available. Third, it requires specialized expertise to capture and annotate the ground truth data for the learning process. Often availability of those experts is limited. Fourth, the dataset is often imbalanced. When we develop an automated disease diagnosis model, we require enough sample data from both the positive and negative class. In medical image analysis, typically more than 90% data belongs to the positive class. Additionally, there are heterogeneous image data available due to different image capturing protocol. Several pre-processing steps are necessary so that the algorithm can work with varying types of data.

Deep Learning models learn automatically from the input data without requiring any hand-crafted feature generation. Convolutional Neural Network (CNN) is the deep learning
model that has achieved tremendous success in image analysis task. In our research work, we are going to use CNN for developing the automated diagnosis model. CNN requires a vast training dataset for efficient training. But in medical image analysis, such large dataset is not available. Designing an efficient classifier that can work with limited training data is a significant challenge in medical image analysis.

Analyzing medical images requires significant expertise. As we are working for Alzheimer’s Disease Diagnosis, we have seen that there is a similarity in the normal healthy brain data of older people and the Alzheimer’s Disease data. It is very challenging to differentiate between both. Diagnosing AD at an early stage is even more challenging, and extensive knowledge and experience are required.

1.9 Contribution

The current thesis presents a multidisciplinary research efforts to investigate the emerging deep learning technologies for medical image analysis. We have developed several models for Alzheimer’s disease diagnosis using deep learning technologies that outperform previous state-of-the-art. We have presented preprocessing approaches for better analysis of MRI data. We have demonstrated approaches that can solve the class-imbalance problem in medical image analysis. We have experimented and introduced several ways to improve the performance of a CNN classifier for disease diagnosis with limited training data. We proposed solution to solve limited dataset problem by generating synthetic medical images and presented approaches to understand the decision of convolutional neural network to discriminate among samples of different disease class.
1.10 List of Publications


Submitted manuscripts


1.11 Thesis Outline

The remainder of this thesis is organized as follows. Chapter 2 presents background study related to deep learning and medical image analysis. Chapter 3 reviews existing approaches to Alzheimer’s Disease Diagnosis and Lung segmentation. Chapter 4, 5, and 6 present our proposed approaches for Alzheimer’s Disease diagnosis. Chapter 7 presents our proposed approach to generate synthetic medical images. Chapter 8 describes our method for understanding behavior of 3D Convolutional Neural Network for Alzheimer’s Disease and Mild Cognitive Impairment diagnosis. Finally, Chapter 9 presents the directions for future work and Chapter 10 concludes the dissertation.
Chapter 2

BACKGROUND STUDY

In this chapter, we discuss an overview of some background ideas of the ensuing sections of this thesis. To set the stage, we review feature learning algorithms and neural network models. Next, we discuss in detail about Convolutional Neural Network (CNN) which is the underlying architecture of the proposed diagnosis frameworks of this thesis. We briefly describe Generative Adversarial Networks that were used to generate synthetic medical images. Finally, we present the theory behind Magnetic Resonance Imaging (MRI), Structural MRI (sMRI) analysis, Positron Emission Tomography (PET), and PET analysis.

Artificial Intelligence (AI) is an area of computer science that focuses on the creation of intelligent machines. AI algorithms simulate intelligent behaviour in computers that can work and behave like a human. Machine Learning is a group of AI techniques that use statistical methods to enable machines to learn with data, without being explicitly programmed and improve with experiences. Deep Learning is a subset of machine learning as shown in Figure 2.1. Deep Learning models solve problems by using neural networks (a series of algorithms modelled after the human brain).

2.1 Learning Algorithms

Machine Learning models are divided into two broad categories - supervised and unsupervised algorithms. In supervised learning, there is input data ($X$) and output variable ($Y$) and the algorithm learns the mapping function from input to output.

$$f : X \rightarrow Y; \quad (2.1)$$

$Y$ typically represents an instance from a fixed set of class. The target of the learning algorithm is to approximate the mapping function well so that when a new input data
Figure (2.1) Relation Among AI, Machine Learning, and Deep Learning.

comes, the model can predict the output class for it. For example, an algorithm trained on the labelled dataset of benign or malignant tumours learns to identify patients with cancer or not.

On the contrary, unsupervised algorithms learn to process data without any label and can identify patterns without any guidance. For example, clustering algorithms can discover the inherent groupings in the data, such as grouping customers by purchasing behaviour.

Figure (2.2) Block Diagram of a Neural Network.
2.2 Neural Network

Neural networks or Artificial neural networks are a set of learning algorithms that replicate the way humans learn. They consist of connected neurons or learning units with some activation function $a$ and parameters $\Theta = \{\omega, \beta\}$, where $\omega$ is a set of weights and $\beta$ is a set of biases. Neural networks are organized in layers consist of neurons. Input data is passed to the network via the ‘input layer’, which communicates to one or more ‘hidden layers’ where the actual processing is done. The hidden layers transform the input and then link to an ‘output layer’ that produces the output. Figure 2.2 shows block diagram of a basic neural network. The activation function is a linear combination of the input to the neurons and the parameters:

$$a = \sigma(\omega^T x + b)$$ (2.2)

Deep learning methods often use neural networks with lots of hidden layers and are often referred to as deep neural networks.

2.3 Convolutional Neural Network

Convolutional neural network (CNN) is a feed-forward artificial neural network and the most popular deep learning model. Artificial neural networks are systems of connected neurons that exchange messages with each other. The connections have respective weights that can be tuned based on experience. As a result, neural networks become adaptive to inputs and capable of learning. Since regular neural networks have fully connected neurons, so they do not scale well to full images. If an image is of size $28 \times 28 \times 3$, a single fully connected neuron in the first hidden layer will have $28 \times 28 \times 3 = 2352$ weights. Now if the image is of size $512 \times 512 \times 3$, the weight would be $512 \times 512 \times 3 = 786,432$ for a single neuron. Now it is very much expected that the full network will have a lot of neurons. As a result, the parameters will add up quickly too. So not only the full connectivity would be wasteful, but also the high number of parameters would lead to overfitting.

CNN is a feed-forward network which means information only moves forward from input
nodes to the output nodes via the hidden nodes (if any). CNN assumes that inputs are images and thus encodes specific properties to the architecture. As a result, the forward function becomes more efficient, and the amount of parameters in the network reduces vastly too.

In a convolutional neural network, the neurons are arranged in three dimensions - weight, height, and depth. So if the input image has size 28x28x3, then the input volume would have dimension 28x28x3. A small region of the layer would be connected to the neurons of the next layer. They won’t be connected fully to all neurons in the next layer. The input image would be reduced to a single vector of class scores by the final output layer. The vector would be arranged along the depth dimension. Each layer converts the 3D input volume to a 3D output volume of neuron activations. The depth of the layer is equal to the number of color channels in the image. Width and heights are determined by the dimension of the image.

Each layer of a CNN uses a differentiable function to convert one volume of activations to another. There are three main types of layers - Convolutional Layer, Pooling Layer, and Fully-Connected Layer. These layers are stacked to form a full CNN architecture. The neurons connected to the local regions of the input computes a dot product of their weights and the respective area. Convolutional layer computes the output of these neurons. Pooling

![Figure (2.3) Illustration of a Convolutional Neural Network.](image-url)
layer performs a downsampling operation along the spatial dimensions - width and height. The fully connected layer where each neuron is connected to all the numbers of the previous volume computes the output class score. Thus, an input image is converted to a final class score by layer by layer transformation of the original pixels. The convolutional and fully connected layer has parameters such as weights and biases of neurons. The transformation of these layers is a function of the activations of the input volume and these parameters. These parameters are trained with gradient descent in such a way so that the output class score is consistent with each training image. The pooling layers have fixed functions. There are two other layers - ReLU layer and Loss layer. ReLU layer increases the nonlinear properties of the decision function. Loss layer uses different loss functions for different tasks. For example, to predict a single class of K mutually exclusive classes, Softmax loss is used.

![Figure (2.4) Local Connectivity of CNN.](image)

### 2.3.1 Local Connectivity

CNN enforces a local connectivity pattern between adjacent layer neurons and thus utilizes spatially-local correlation. Figure 2.4 shows local connectivity in a CNN. Here the inputs of hidden units of layer \( m \) are generated from a subset of units of layer \( m-1 \). The units of layer \( m-1 \) have spatially adjacent receptive fields. Let us consider layer \( m-1 \) as input. Layer \( m \) units are connected to 3 adjacent neurons in the layer. So we can say that the receptive field width of layer \( m \) is 3. Receptive field width of layer \( m+1 \) is also 3 concerning layer \( m \). But with respect to the input layer, the receptive field width of layer \( m+1 \) is 5. Each unit does not respond to variations outside of its receptive field with respect to the input layer. So it is ensured that the learned filters will produce the strongest response to a spatially
local input pattern. Now if we stack multiple layers, then the filters would become global gradually and response to a larger region of pixel space. As in the figure, we can see neurons of hidden layer m+1 can encode a feature of width 5.

2.3.2 Shared Weights

One crucial feature of CNN is each filter $h_i$ is replicated across the entire visual field. Same weight and bias are shared by these replicated units, and a feature map is produced. By sharing the same weights, CNN achieves shift-invariant property. So CNN can detect features regardless of their position. Sharing weight thus allows CNN to achieve high performance in recognition and detection problem. The need to learn free parameters is also decreased significantly, so learning efficiency improves. It also reduces the required memory size. Figure 2.5 shows a feature map consisted of 3 units. We obtain a feature map by repeatedly applying a function across the sub-regions of the input image that is the convolution of the image with a linear filter and adding a bias term and applying a non-linear function, $\sigma$. The $k$–th feature map at a given layer can be denoted as $h^k$. If the filters of feature map $h^k$ is determined by weights $W^k$ and bias $b_k$, then we can write:

$$h^k_{ij} = \sigma((W^k * x)_{ij} + b_k)$$

(2.3)

2.3.3 Pooling

Pooling operation reduces the spatial size of the data representation and helps to reduce the parameters and computation in the network. Max-pooling and average pooling are
two popular pooling operations. Max pooling is also known as down-sampling. The input image is partitioned into a set of non-overlapping rectangles. Then for each sub-region, the maximum value is computed. Since non-maximal values are eliminated, so computation for upper layers are reduced. Max pooling also provides translation invariance. If we cascade a max-pooling layer with a convolutional layer, we get eight directions to translate the input image by a single pixel. Now if max-pooling is done in a 2x2 region, out of the eight possible configurations, three will provide the same output at the convolutional layer. If the region size is 3x3, 5 out of 8 regions will give the same output. Because of this robustness, max-pooling is used to reduce the dimension of intermediate representations. Figure 2.6 presents an example of max-pooling operation.

![Max-Pooling Operation](image)

Figure (2.6) Max-Pooling Operation. Each Max is Taken Over 4 Numbers Arranged in 2x2 Square.

### 2.4 Generative Adversarial Networks

Generative Adversarial Networks (GANs) is a deep learning architecture that consisted of two models - a generative model G and a discriminative model D. The generative model captures the data distribution. The discriminative model estimates the probability that the sample is drawn from the training data rather than the generative model. The two models are simultaneously trained via an adversarial process. The architecture is inspired by game theory and corresponds to a minimax two-player game. The training procedure of G is to maximize the probability of D making a mistake [6].

Let the generator $G(z, \theta_x)$ is a differentiable function represented by a multilayer percep-
tron with parameters $\theta_g$ that depicts a mapping to the data space. To learn the generator’s
distribution $\rho_g$ over the data space $x$, a prior $\rho_z$ is defined on random input noise variables
$z$. The discriminator $D(x, \theta_d)$ is also a neural network that gets a sample the real dataset
or the generated synthetic dataset produced by G and outputs a single scalar value that
the input data comes from the real training dataset. The training process focuses on the
task that the discriminator D will maximize the probability of assigning correct labels to the
training examples and generated samples from G. At the same time, G is trained to generate
data samples similar to the real dataset so that D cannot differentiate them from actual
data. Similar to game theory, the discriminator D and the generator G play a two-player
mini-max game with following value function $V(G, D)$:

$$
\min_G \max_D V(D, G) = \mathbb{E}_{x \sim \rho_{data}}[\log D(x)] + \mathbb{E}_{z \sim \rho_{data}}[\log (1 - D(z))]
$$

(2.4)

Where $x$ is the real data and $z$ is the input random noise. $\rho_{data}$, $\rho_z$ represent the distri-
bution of the real data and the input noise. $D(x)$ represents the probability that $x$ came from
the real data while $G(z)$ represents the mapping to synthesize the real data. The generator,
G is a deeper neural network and have more convolutional layers and nonlinearities. The
noise vector $z$ is upsampled while G learns the weights through backpropagation. At some
point, the generator starts producing data that is classified as real by the discriminator.

Figure (2.7) The Major Components of A Magnetic Resonance Imaging System [1].
2.5 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a technique that creates a 3D representation of body structures using magnetic fields and radio waves. Nowadays it is a standard practice to use MRI to detect changes in the body caused by different diseases. Magnetic Resonance Imaging was developed around 1980. MRI is a powerful imaging technique to visualizing detailed structures in vivo. MRI technique uses magnetic manipulation of protons to acquire images without ionizing radiation. In the MRI scanner, the patient is placed in a strong magnetic field that causes the hydrogen atoms (protons in the water molecules) in the patients body to align either in parallel or anti-parallel to the magnetic field.

Figure 2.7 shows the major components of a Magnetic Resonance Imaging system. Protons are randomly oriented within the water nuclei of the tissue. Radio-frequency (RF) pulses emitted from the radio-frequency coils in the MRI system causes the proton to spin on its axis. When the RF pulse is turned off, protons return to their resting alignment and emit RF energy. The emitted signals are measured after a certain period following the initial RF to produce the Three Dimensional (3D) grey-scale image. Proton spin relaxation rate differs depending on their tissue type and decides the intensity level of the image. Different types of images are created using a different sequence of RF pulses. Repetition Time (TR) refers to the amount of time between consecutive pulse sequences applied to the same slice. Time to Echo (TE) indicates the time between the transmission of the RF pulse and the receipt of the echo signal [1].

Figure (2.8) T1 and T2 Weighted Images of Brain [2].
T1 and T2 are two commonly used relaxation times for protons. T1 or longitudinal relaxation time is the measure of time taken for spinning protons to realign with the external magnetic field. It determines the rate at which excited protons return to equilibrium. T2 or transverse relaxation time is the measure of the time taken for spinning protons to lose phase coherence among the nuclei spinning perpendicular to the main field. It determines the rate at which excited protons reach equilibrium or go out of phase with each other [2]. Figure 2.8 shows T1 and T2 weighted images of brain. T1-Weighted MRI is produced with short TE and short TR (TR <1000 ms, TE <30 ms ) while long TE and long TR (TR >2000 ms, TE >80 ms) is used to obtain T2-weighted images. In a T1-weighted brain MRI, the Gray Matter (GM) is visible as a dark grey area, the White Matter (WM) is light grey, and the Cerebrospinal Fluid (CSF) is black. On the other hand, in T2-weighted brain images, GM is light grey, WM is dark grey, and CSF is bright.

![Figure (2.9) Structural MRI Images Presenting (a) Normal Control; (b) MCI; (c) AD.](image)

2.5.1 Structural MRI (sMRI)

Structural magnetic resonance imaging (sMRI) is used to examine the physical structure of the brain. It translates the differences in the water content with different shades of grey and outlines the shape and size of the subregions of the brain. Structural MRI presents good
tissue contrast that helps to identify changes in the brain such as the presence of tumours. For our research, we analyze sMRI for Alzheimer’s Disease diagnosis. Figure 2.9 shows structural MRI images presenting different stages of Alzheimer’s Disease.

### 2.6 Positron Emission Tomography (PET)

Positron Emission Tomography is a class of nuclear medicine imaging. It is also known as PET imaging or a PET scan. Nuclear medicine refers to a type of medical imaging methods that utilizes radioactive material in a small amount to diagnose and determine the stage of a disease or treat a disease. Nuclear Medicine procedures are known for their ability to pinpoint molecular activity within the body. So, they have the potential to diagnose the disease in the earliest stage and find patient’s response to therapeutic interventions. These procedures are used to diagnose different types of cancers, heart disease, gastrointestinal, endocrine, neurological disorders and other abnormalities within the body. Radioactive materials known as radiopharmaceuticals or radiotracers are used to perform nuclear medicine imaging procedures.

Positron Emission Tomography (PET) uses small amounts of radiotracers, a special cam-

![Figure (2.10) Positron Emission Tomography (PET) Scanning Procedure [3].](image)

era and a computer to evaluate the physiological changes. Positron Emission Tomography
(PET) measures the body changes at the cellular level by looking at blood flow, metabolism, neurotransmitters, and radiolabelled drugs. PET may identify early onset of disease before it is evident on other imaging tests. It performs quantitative analyses and finds relative changes over time as the disease process evolves or in response to a specific stimulus [3]. Figure 2.10 presents the procedure to perform PET scan. Depending on the organ or tissue to study, the tracer can be injected, swallowed or inhaled. A small amount of a radioactive tracer is administered into a peripheral vein, and the radioactivity emitted is measured. The tracer is injected as an intravenous injection usually labelled with oxygen-15, fluorine-18, carbon-11, or nitrogen-13. The areas of the disease often demonstrate higher levels of chemical activity and the tracer collects in areas of your body. On the PET scan, these areas might show up as bright spots. The total radioactive dose is similar to the dose used in Computed Tomography (CT).

![PET Images](image)

Figure (2.11) Example of Brain PET Images (a) Normal Control (b) MCI (c) AD.

PET imaging tracers can correlate β-amyloid deposition in the brain. The amyloid deposition in the brain can be detected years before the onset of clinical signs of Alzheimer’s Disease. The PET imaging tracers can help in differentiating dementia syndromes at the early stage. PET scans reflect the resting state cerebral metabolic rates of glucose that is an indicator of neuronal activity. Cerebral glucose metabolic alterations have distinct patterns that can be used to identify Alzheimer’s Disease symptoms. Fig. 2.11 shows some brain PET images with different AD stages.
Chapter 3

REVIEW OF THE STATE-OF-THE-ART

In this section, we review the related work focusing on Alzheimer’s Disease Diagnosis, synthetic medical image generation and CNN Visualization.

3.1 Alzheimer’s Disease Diagnosis

Detection of physical changes in brain complement clinical assessments and has an increasingly important role for early detection of AD. Researchers have been devoting their efforts to Neuroimaging techniques to measure pathological brain changes related to Alzheimer’s disease. Machine learning techniques have been developed to build classifiers using imaging data and clinical measures for AD diagnosis [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. These studies have identified the significant structural differences in the regions such as the hippocampus and entorhinal cortex between the healthy brain and brain with AD. Changes in cerebrospinal tissues can explain the variations in the behavior of the AD patients [17], [18]. Besides, there is a significant connection between the changes in brain tissues connectivity and behavior of AD patient [19]. The changes causing AD due to the degeneration of brain cells are noticeable on images from different imaging modalities, e.g., structural and functional Magnetic Resonance Imaging (sMRI, fMRI), Position Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), and Diffusion Tensor Imaging (DTI) scans. Several researchers have used these neuroimaging techniques for AD Diagnosis. For example, sMRI ([20], [21], [22], [23], [24], [25]), fMRI([26], [27]), PET ([28], [29]), SPECT ([30], [31], [32]), and DTI ([33], [34]) have been used for diagnosis or prognosis of AD. Moreover, information from multiple modalities have been combined to improve the diagnosis performance ([35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46]).
A classic Magnetic Resonance Imaging (MRI) based automated AD diagnostic system has mainly two building blocks - feature/biomarker extraction from the MRI data and classifier based on those features/biomarkers. Though various types of feature extraction techniques exist, there are three major categories - (1) voxel-based approach, (2) Region of Interest (ROI)-based approach, and (3) patch-based approach. Voxel-based approaches are independent of any hypothesis on brain structures [47], [48], [49], [50]. For example, voxel-based morphometry measures local tissue (i.e., white matter, gray matter, and cerebrospinal fluid) density of the brain. Voxel-based approaches exploit the voxel intensities as the classification feature. The interpretation of the results are simple and intuitive in voxel-based representations, but they suffer from the over-fitting problem. Since there are limited (e.g., tens or hundreds) subjects with very high (millions) dimensional features [51], which is a major challenge for AD diagnosis based on neuroimaging. To achieve more compact and useful features, dimensionality reduction is essential. Moreover, voxel-based approaches suffer from the ignorance of regional information.

Region of Interest (ROI)-based approach utilizes the structurally or functionally pre-defined brain regions and extracts representative features from each region [21], [24], [27], [28], [52], [53],[54]. These studies are based on specific hypothesis on abnormal regions of the brain. For example, some studies have adopted gray matter volume [55], hippocampal volume [56], [57], [58] and cortical thickness [21], [59]. ROI-based approaches are widely used due to relatively low feature dimensionality and whole brain coverage. But in ROI-based approaches, the extracted features are coarse as they cannot represent small or subtle changes related to brain diseases. The structural or functional changes that occur in the brain due to neurological disorder are typically spread to multiple regions of the brain. As the abnormal areas can be part of a single ROI or can span over multiple ROIs, voxel-based or ROI-based approaches may not efficiently capture the disease-related pathologies. Besides, the Region of Interest (ROI) definition requires expert human knowledge. Patch-based approaches [20], [60], [61], [62], [63], [64], [65] divide the whole brain image into small-sized patches and extract feature vector from those patches. Patch extraction does not require ROI
identification, so the necessity of human expert involvement is reduced compared to ROI-based approaches. Compared to voxel-based approaches, patch-based methods can capture the subtle brain changes with significantly reduced dimensionality. Patch-based approaches learn from the whole brain and better captures the disease related pathologies that results in superior diagnosis performance. However, there is still challenges to select informative patches from the MRI images and generate discriminative features from those patches.

A large number of research works focused on developing advanced machine learning models for AD diagnosis using MRI data. Support Vector Machine SVM), Logistic Regressors (e.g., Lasso, and Elastic Net), Sparse Representation based Classification (SRC), Random Forest Classifier, etc. are some widely used approaches. For example, Kloppel et al. [50] used linear SVM to detect AD patients using T1 weighted MRI scan. Dimensional reduction and variations methods were used by Aversen et al. [66] to analyze structural MRI data. They have used both SVM binary classifier and multi-class classifier to detect AD from MRI images. Vemuri et al. [67] used SVM to develop three separate classifiers with MRI, demographic and genotype data to classify AD and healthy patients. Katherine Gray [68] developed a multi-modal classification model using random forest classifier for AD diagnosis from MRI and PET data. Amulya et al. [69] used Gray-Level Co-occurrence Matrix (GLCM) method for AD classification. Morra et al. [70] compared several model’s performances for AD detection including hierarchical AdaBoost, SVM with manual feature and SVM with automated feature. For developing these classifiers, typically predefined features are extracted from the MRI data. However, training a classifier independent from the feature extraction process may result in sub-optimal performance due to the possible heterogeneous nature of the classifier and features [71].

Recently, deep learning models have been famous for their ability to learn feature representations from the input data. Deep learning networks use a layered, hierarchical structure to learn increasingly abstract feature representations from the data. Deep learning architectures learn simple, low-level features from the data and build complex high-level features in a hierarchy fashion. Deep learning technologies have demonstrated revolutionary performance
in several areas, e.g., visual object recognition, human action recognition, natural language processing, object tracking, image restoration, denoising, segmentation tasks, audio classification, brain-computer interaction, etc. In recent years, deep learning models specially Convolutional Neural Network (CNN) have demonstrated excellent performance in the field of medical imaging, i.e., segmentation, detection, registration, and classification [72]. For neuroimaging data, deep learning models can discover the latent or hidden representation and efficiently capture the disease-related pathologies. So, recently researchers have started using deep learning models for AD and other brain disease diagnosis.

Gupta et al. [64] have developed a sparse autoencoder model for AD, Mild Cognitive Impairment (MCI) and healthy control (HC) classification. Payan et al. [65] trained sparse autoencoders and 3D CNN model for AD diagnosis. They also developed a 2D CNN model that demonstrated nearly identical performance. Brosch et al. [73] developed a deep belief network model and used manifold learning for AD detection from MRI images. Hosseini-As et al. [74] adapted a 3D CNN model for AD diagnostics. Liu et al. [75] developed a deep learning model using both unsupervised and supervised techniques and classified AD and MCI patients. Liu et al. [76] have developed a multimodal stacked auto-encoder network using zero-masking strategy. Their target was to prevent loss of any information of the image data. They have used SVM to classify the neuroimaging features obtained from MR/PET data. Sarraf et al. [77] used fMRI data and deep LeNet model for AD detection. Suk et al. [20], [41], [78], [79] developed an autoencoder network-based model for AD diagnosis and used several complex SVM kernels for classification. They have extracted low to mid level features from magnetic current imaging (MCI), MCI-converter structural MRI, PET data and performed classification using multi-kernel SVM. Cárdenas-Peña et al. [80] have developed a deep learning model using central kernel alignment and compared the supervised pre-training approach to two unsupervised initialization methods, autoencoders and Principal Component Analysis (PCA). Their experiment shows that SAE with PCA outperforms three hidden layers SAE and achieves an increase of 16.2% in overall classification accuracy.

So far, AD is detected at a much later stage when treatment can only slow the pro-
gression of cognitive decline. No treatment can stop or reverse the progression of AD. So, early diagnosis of AD is essential for preventive and disease-modifying therapies. Most of the existing research work on AD diagnosis focused on binary classification problems, i.e., differentiating AD patients from healthy older adults. However, for early diagnosis, we need to distinguish among current AD stages, which makes it a multi-class classification problem. In our initial work [81], we developed a very deep convolutional network and classified the four different stages of the AD - nondemented, very mild dementia, mild dementia, and moderate dementia. We improved the previous model, developed an ensemble of deep convolutional neural networks [82], and demonstrated better performance on the Open Access Series of Imaging Studies (OASIS) dataset [83]. Additionally, we developed an efficient deep convolutional neural network based classifier [84] and demonstrated better performance on the ADNI dataset [85].

3.2 Medical Image Synthesis

Medical Image synthesis and Generative Adversarial networks have got attention in recent years. Costa et al. [86] used a fully-convolutional neural network to train on retinal vessel segmentation images and then applied GANs for generating synthetic retinal images. Dai et al. [87] used GANs for creating lung fields and heart segmentation images from chest X-ray images. Shin et al. [88] utilized GANs for generating synthetic abnormal MRI images with brain tumors. Nie et al. [89], proposed an auto-context model for brain CT and MRI image refinement. Schlegl et al. [90] trained GANs for anomaly detection in retinal images. Frid-Adar et al. [91] applied GANs for synthesizing liver lesion ROIs to apply in liver lesion classification. Hu et al. [92] applied GANs to generate a MRI motion model. Mahapatra et al. [93] synthesized high-resolution retinal fundus images using generative adversarial networks. Nie et al. [89] generated synthetic pelvic CT images using GANs.

In our previous research works, we had to handle the limited dataset problem for Alzheimer’s Disease diagnosis. There is a gap in research work for synthesizing brain images
for Alzheimer’s Disease diagnosis. Besides, there are very few works done for PET image synthesis. To mitigate these gaps, we propose a novel model to generate synthetic brain Positron Emission Tomography (PET) images exploiting Generative Adversarial Networks for three stages of Alzheimer’s Disease - Normal Control (CN), Mild Cognitive Impairment (MCI), and Alzheimer’s Disease (AD).

### 3.3 CNN Visualization

Recently, several visualization methods for CNN interpretation have been proposed. Some of these methods focus on synthesizing the image that maximizes the score of a given unit in a pretrained CNN, while others invert feature maps of a conv-layer back to the input image. Among the gradient-based methods, deconvnet with Occlusion Sensitivity [94], Sensitivity analysis with backpropagation [95], Guided Backpropagation (GB) [96] demonstrated notable performance. These methods compute the gradients of the score of the convolutional neural network with respect to the input image. The gradients that maximize the neuron score are used to estimate the image appearance.

Rieke et al. [97] compared four visualization methods to understand a CNN behavior for AD diagnosis using MRI data. Böhole et al. [98] presented an approach utilizing Layerwise Relevance Propagation (LRP) [99] for CNN visualization for AD diagnosis. Korolev et al. [100] applied the occlusion method to interpret a deep CNN. They mainly focused on developing a 3D-CNN classifier and didn’t employ different visualization method. Yang et al. [11] presented segmentation-based occlusion approach to find visual explanations that can indicate a 3D-CNNs spatial attention on MRI brain scans when making predictions. To the best of our knowledge, our work is the first to understand CNN behavior for Alzheimer’s Disease and Mild Cognitive Impairment diagnosis trained using PET data.
Chapter 4

A NOVEL DEEP LEARNING BASED MULTI-CLASS CLASSIFICATION METHOD FOR ALZHEIMER’S DISEASE DIAGNOSIS

4.1 Introduction

In this chapter, we describe our initial work and present a novel deep learning model for multi-Class Alzheimer’s Disease detection and classification using Brain MRI Data. We design a very deep convolutional network and demonstrate the performance on the Open Access Series of Imaging Studies (OASIS) dataset [83]. Fig. 7.1 shows some brain MRI images presenting different AD stage.

![Example of Different Brain MRI Images Presenting Different AD Stage](image)

Figure (4.1) Example of Different Brain MRI Images Presenting Different AD Stage. (a) Nondemented; (b) very mild dementia; (c) mild dementia; (d) moderate dementia.

4.2 Method

In this section, the proposed Alzheimer’s disease detection and classification framework would be presented. The proposed model is shown in Fig. 4.2. Our model is inspired by Inception-V4 network [101]. After the preprocessing is done, the input is passed through a stem layer. A stem layer includes several 3*3 convolution layers, 1*1 convolution layer, and max pooling layer. There is seven 3*3 convolution layer connected in different stages
and two filter-expansion layers (1*1 convolution layer). Inception-A module has four filter-expansion layers, three 3*3 convolution layer, and one average pooling layer. Inception-B module has four filter-expansion layers, four 1*7 convolution layer, two 7*1 convolution layer and one average pooling layer. Inception-C module has four filter-expansion layers, three 1*3 convolution layer, three 3*1 convolution layer and one average pooling layer. Reduction-A module has one filter-expansion layer, three 3*3 convolution layer, and one 3*3 max-pooling layer. The Reduction-B module has two filter-expansion layers, two 3*3 convolution layer, one 1*7 convolution layer, one 7*1 convolution layer and one 3*3 max pooling layer. The input and output of all these modules pass through filter concatenation process. We have redesigned the final softmax layer for Alzheimer’s disease detection and classification. The softmax layer has four different output class: nondemented, very mild, mild and moderate AD. The network takes an MRI image as input and extracts layer-wise feature representation from the first stem layer to the last drop-out layer. Based on this feature representation, the input MRI image is classified to any of the four output classes.

Figure (4.2) Block Diagram of Proposed Alzheimer’s Disease Detection and Classification Framework.
To measure the loss of the proposed network, we have used cross entropy. The Softmax layer takes the feature representation, \( f_i \) and interprets it to the output class. A probability score, \( p_i \) is also assigned for the output class. If we define the number of Alzheimer’s disease stages as \( m \), then we get

\[
p_i = \frac{\exp(f_i)}{\sum_i \exp(f_i)}, i = 1, ..., m
\]

and

\[
L = - \sum_i t_i \log(p_i),
\]

where \( L \) is the loss of cross entropy of the network. Back propagation is used to calculate the gradients of the network. If the ground truth of an MRI image is denoted as \( t_i \), then,

\[
\frac{\partial L}{\partial f_i} = p_i - t_i
\]

There is numerous possible combination for the hyper-parameters of a network. It takes a lot of time and effort to decide a stable hyperparameter set for a network. To reduce this time, we have used hyperparameters of the Inception-V4 model [101] instead of random initialization. The weights and biases of the inception-v4 model [101] pre-trained with ImageNet database [102] provide our network an efficient hyperparameter set. As a result, the model has a sense of better feature detector and can use that knowledge for learning features from the small medical image dataset. We have trained our model with OASIS [83] dataset. To prevent overfitting in the network, we have applied data augmentation technique such as reflection and scaling.

4.3 Data

OASIS dataset is prepared by Dr. Randy Buckner from the Howard Hughes Medical Institute (HHMI) at Harvard University, the Neuroinformatics Research Group (NRG) at Washington University School of Medicine, and the Biomedical Informatics Research Network (BIRN) [83]. There are 416 subjects aged 18 to 96, and for each of them, 3 or 4
T1-weighted MRI scans are available. 100 of the patients having age over 60 are included in the dataset with very mild to moderate AD. Fig. 4.3 shows some sample brain MRI images from OASIS dataset.

Figure (4.3) Sample Images From OASIS Dataset.

4.4 Experiments

We have implemented the proposed deep CNN model for Alzheimer’s disease detection and classification using Tensorflow [30] and Python on a Linux X86-64 machine with AMD A8 CPU, 16 GB RAM and NVIDIA GeForce GTX 770. We have applied data augmentations techniques - scaling and reflection on the images. Since the dataset is small, 5-fold cross validation is performed on the dataset. For each fold, We have used 70% as training data, 10% as validation data and 20% as test data. The input size of the Inception-V4 network [101] is 299*299*3. To fit the MRI data, we have designed the input size of our network as 299*299*1. We have modified the Inception B and C module so that they can accept the MRI data. The convolutional filter size of Inception-B is 1154 in the original network. We made it to 1152 to fit the MRI data. The convolutional filter size of Inception-C is 2048 in the original network. We made it to 2144 to fit the MRI data. The network is optimized with the RMSProp [103] algorithm and early-stopping is used for regularization. The decay of the network is 0.9 and batch size is 8. The base learning rate is set to 0.045.
Table (4.1) Confusion Matrix

<table>
<thead>
<tr>
<th>AD stage</th>
<th>Nondemented</th>
<th>Very mild</th>
<th>Mild</th>
<th>Moderate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nondemented</td>
<td>52</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Very mild</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mild</td>
<td>7</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Moderate</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

4.4.1 Results

To our best knowledge, our approach is the first one for Alzheimer’s disease detection and classification using deep learning method on OASIS dataset. So, we are not comparing it with previous traditional methods. The current accuracy of our method is 73.75%. The confusion matrix is presented in Table 4.1. The proposed model is much faster and takes less than 1 hour to train and test the OASIS dataset for Alzheimer’s disease detection and classification. This performance is superior than all previous traditional methods. It would take weeks for human experts to analyze and classify all the MRI data. We do not need any manual hand-crafting for feature generation in our model.

We have implemented another deep model with traditional inception module and 22 layers following GoogleNet [104] architecture and compared the performance with our proposed model. The performance comparison is presented in Table 4.2.

Table (4.2) Five-fold Cross Validation Performance Accuracy comparison.

<table>
<thead>
<tr>
<th>No. of epochs</th>
<th>Traditional Inception Network</th>
<th>Proposed Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>60.00%</td>
<td>71.25%</td>
</tr>
<tr>
<td>10</td>
<td>64.25%</td>
<td>73.75%</td>
</tr>
</tbody>
</table>

4.5 Summary

An automated Alzheimer’s disease detection and classification framework is crucial for the early diagnosis and treatment of the AD patients. Our proposed model is faster, does not need any handcrafted feature, and can handle the small medical image dataset. We have
provided a one-step analysis for the brain MRI data for AD detection and classification.
Chapter 5

ALZHEIMER'S DISEASE DIAGNOSIS USING AN ENSEMBLE SYSTEM OF DEEP CONVOLUTIONAL NEURAL NETWORKS

5.1 Introduction

In this chapter, we improve our work presented in Part 4 and present an ensemble of deep Convolutional Neural Networks for Alzheimer’s Disease diagnosis and demonstrate superior performance on the Open Access Series of Imaging Studies (OASIS) dataset. While most of the existing approaches perform binary classification, our model can identify different stages of Alzheimer's Disease.

Machine learning studies using neuroimaging data for developing diagnostic tools helped a lot for automated brain MRI segmentation and classification. Most of them use handcrafted feature generation and extraction from the MRI data. These hand-crafted features are fed into machine learning models such as Support Vector Machine, Logistic regression model, etc. for further analysis. Human experts play a crucial role in these complex multi-step architectures. Moreover, neuroimaging studies often have a dataset with limited samples. While image classification datasets used for object detection and classification have millions of images (for example, ImageNet database [105]), neuroimaging datasets usually contain a few hundred images. But a large data set is vital to develop robust neural networks. Because of the scarcity of large image database, it is important to develop models that can learn useful features from the small dataset. Moreover, The state-of-the-art deep learning models are optimized to work with natural (every day) images. These models also require a lot of balanced training data to prevent over-fitting in the network. We developed a deep convolutional neural network that learned features directly from the input sMRI and eliminated the need for the hand-crafted feature generation. We trained our model using the OASIS database [83] that has only 416 sMRI data. Our proposed model can classify different
stages of Alzheimer’s Disease and outperforms the off-the-shelf deep learning models.

### 5.2 Method

Let \( x = \{x_i, \ i = 1, ..., N\} \), a set of MRI data with \( x_i \in [0, 1, 2, ..., L - 1]^{h \times w \times l} \), a three Dimensional (3D) image with \( L \) gray scale values, \( h \times w \times l \) voxels and \( y \in \{0, 1, 2, 3\} \), one of the stages of AD where 0, 1, 2, 3 refers to nondemented, very mild dementia, mild dementia, and moderate dementia respectively. We will construct a classifier,

\[
f : X \rightarrow Y; \ x \mapsto y
\]

(5.1)

that predicts a label \( y \) in response to an input image \( x \) with minimum error rate. Mainly, we want to determine this classifier function \( f \) by an optimal set of parameters \( w \in \mathbb{R}^P \) (where \( P \) can easily be in the tens of millions), that will minimize the loss or error rate of prediction. The training process of the classifier would be an iterative process to find the set of parameters \( w \), that minimizes the classifier’s loss

\[
L(w, X) = \frac{1}{n} \sum_{i=1}^{n} l(f(x_i, w), \hat{c}_i)
\]

(5.2)

where \( x_i \) is \( i^{th} \) image of \( X \), \( f(x_i, w) \) is the classifier function that predicts the class \( c_i \) of \( x_i \) given \( w \), \( \hat{c}_i \) is the ground-truth class for \( i^{th} \) image \( x_i \) and \( l(c_i, \hat{c}_i) \) is the penalty function for predicting \( c_i \) instead of \( \hat{c}_i \). We set \( l \) to the loss of cross–entropy,

\[
l = - \sum_{i} \hat{c}_i \log c_i
\]

(5.3)

#### 5.2.1 Data Selection

In this study, we use the OASIS dataset [83] prepared by Dr. Randy Buckner from the Howard Hughes Medical Institute (HHMI) at Harvard University, the Neuroinformatics Research Group (NRG) at Washington University School of Medicine, and the Biomedical
Informatics Research Network (BIRN). There are 416 subjects aged 18 to 96, and for each of them, 3 or 4 T1-weighted sMRI scans are available. 100 of the patients having age over 60 are included in the dataset with very mild to moderate AD.

5.2.2 Data Augmentation

Data augmentation refers to artificially enlarging the dataset using class-preserving perturbations of individual data to reduce the overfitting in neural network training [106]. The reproducible perturbations will enable new sample generation without changing the semantic meaning of the image. Since manually sourcing of additional labeled image is difficult in medical domain due to limited expert knowledge availability, data augmentation is a reliable way to increase the size of the dataset. For our work, we developed an augmentation scheme involving cropping for each image. We set the dimension of the crop similar to the dimension of the proposed deep CNN classifier. Then, we extracted three crops from each image, each for one of the image plane: Axial or horizontal plane, Coronal or frontal plane, and Sagittal or median plane. For our work, we use 80% data from the OASIS dataset as training set, and 20% as test dataset. From the training dataset, a random selection of 10% images is used as validation dataset. The augmentation process is performed separately for the train, validation and test dataset. One important thing to consider is the data augmentation process is different from classic cross-validation scheme. Data augmentation is used to reduce overfitting in a vast neural network while training with a small dataset. On the other-hand, cross-validation is used to derive a more accurate estimate of model prediction performance. Cross-validation technique is computationally expensive for a deep convolutional neural network training as it takes an extensive amount of time.

5.2.3 Network Architecture

Our proposed network is an ensemble of three deep convolutional neural networks with slightly different configuration. We made a considerable amount of effort for the design of the proposed system and the choice of the architecture. All the individual models have a
common architectural pattern consisted of four basic operations:

- convolution
- batch normalization [107]
- rectified linear unit, and
- pooling
as dense connectivity [108] as shown in Fig. 5.1. The dense connections have a regularizing effect that reduces overfitting in the network while training with a small dataset. We keep these layers very narrow (e.g., 12 filters per layer) and connect each layer to every other layer. Similar to [108], we will refer to the layers as dense layer and combination of the layers as dense block. Since all the dense layers are connected to each other, the $i^{th}$ layer receives the feature-maps $(h_0, h_1, h_2, ..., h_{i-1})$, from all previous layers $(0, 1, 2, ..., i - 1)$. Consequently, the network has a global feature map set, where each layer adds a small set of feature-maps. In times of training, each layer can access the gradients from the loss function as well as the original input. Therefore, the flow of information improves, and gradient flow becomes stronger in the network. Fig. 5.2 shows the intermediate connection between two dense
blocks.

For the design of the proposed system, we experimented with several different deep learning architectures and finally developed an ensemble of three homogeneous deep convolution neural networks. The proposed model is shown in Fig. 5.3. We will refer to the individual models as $M_1$, $M_2$, and $M_3$. In Fig. 5.3, the top network is $M_1$, the middle network is $M_2$ and the bottom network is $M_3$. Each of the models consists of several convolution layers, pooling layers, dense blocks and transition layers. The transition layer is a combination of batch normalization layer, a 1*1 convolutional layer followed by a 2*2 average pooling layer with stride 2. Batch normalization [107] acts as a regularizer and speeds up the training process dramatically. Traditional normalization process (shifting inputs to zero-mean and unit variance) is used as a pre-processing step. Normalization is applied to make the data comparable across features. When the data flow inside the network at the time of training process, the weights and parameters are continuously adjusted. Sometimes these adjustments make the data too big or too small; a problem referred as 'Internal Covariance Shift'. Batch normalization largely eliminates this problem. Instead of doing the normalization at the beginning, batch normalization is performed to each mini batches along with SGD training. If $\mathcal{B} = \{x_1, x_2, ..., x_m\}$ is a mini-batch of $m$ activations value, the normalized values are $(\hat{x}_1, \hat{x}_2, ..., \hat{x}_m)$ and the linear transformations are $y_1, y_2, ..., y_m$, then batch normalization is referred to the transform:

$$BN_{\gamma, \beta} : x_1, x_2, ..., x_m \rightarrow y_1, y_2, ..., y_m \quad (5.4)$$

Considering $\gamma, \beta$ the parameters to be learned and $\epsilon$, a constant added to the mini-batch
variance for numerical stability, batch normalization is given by the following equations:

\[
\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^{m} x_i \\
\sigma^2_B \leftarrow \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2 \\
\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma^2_B + \epsilon}} \\
y_i \leftarrow \gamma \hat{x}_i + \beta \equiv BN_{\gamma, \beta}(x_i)
\]

where \( \mu_B \) is mini-batch mean, \( \sigma^2_B \) is mini-batch variance [107].

Though each model has four dense blocks, they differ in the number of their internal 1*1 convolution and 3*3 convolution layers. The first model, \( M_1 \), has six (1*1 convolution and 3*3 convolution layers) in the first dense block, twelve (1*1 convolution and 3*3 convolution layers) in the second dense block, twenty-four (1*1 convolution and 3*3 convolution layers) in the third dense block and sixteen (1*1 convolution and 3*3 convolution layers) in the fourth dense block. The second model, \( M_2 \), and third model, \( M_3 \), has (6, 12, 32, 32) and (6, 12, 36, 24) arrangement respectively. Because of the dense connectivity, each layer has direct connections to all subsequent layers, and they receive the feature-maps from all preceding layers. So, the feature-maps work as global state of the network, where each layer can add their own feature-map. The global state can be accessed from any part of the network and how much each layer can contribute to is decided by the growth rate of the network. Since the feature-maps of different layers are concatenated together, the variation in the input of subsequent layers increases and results in more efficiency.

The input MRI is 3D data, and our proposed model is a 2D architecture, so we devise an approach to convert the input data to 2D images. For each MRI data, we created patches from three physical planes of imaging: Axial or horizontal plane, Coronal or frontal plane, and Sagittal or median plane. These patches are fed to the proposed network as input. Besides, this data augmentation technique increases the number of samples in training data set. The size of each patch is 112*112. We trained the individual models separately and
each of them has own Softmax layer for classification decision. The Softmax layers have four
different output classes: nondemented, very mild, mild, and moderate AD. The individual
models take the input image and generate its learned representation. The input image is
classified to any of the four output classes based on this feature representation. To measure
the loss of each of these models, we used cross entropy. The Softmax layer takes the learned
representation, \( f_i \) and interprets it to the output class. A probability score, \( p_i \) is also assigned
for the output class. If we define the number of output classes as \( m \), then we get

\[
p_i = \frac{\exp(f_i)}{\sum_i \exp(f_i)}, \quad i = 1, ..., m
\]

and

\[
L = -\sum_i t_i \log(p_i)
\]

where \( L \) is the loss of cross entropy of the network. Back propagation is used to calculate
the gradients of the network. If the ground truth of an MRI data is denoted as \( t_i \), then,

\[
\frac{\partial L}{\partial f_i} = p_i - t_i
\]

To handle the imbalance in the dataset, we used cost sensitive training [109]. A cost matrix
\( \xi \) was used to modify the output of the last layer of the individual networks. Since the less
frequent classes (very mild dementia, mild dementia, moderate dementia) are underrepre-
sentated in the training dataset, the output of the networks were modified using the cost
matrix \( \xi \) to give more importance to these classes. If \( o \) is the output of the individual model,
\( p \) is the desired class and \( L \) is the loss function, then \( y \) denotes the modified output:

\[
y^i = L(\xi_p, o^i) : y^i_p \geq y^i_j \forall j \neq p
\]
The loss function is modified as:

\[ L = -\sum_n t_n \log(y_n) \]  

(5.10)

where \( y_n \) incorporates the class-dependent cost \( \xi \) and is related to the output \( o_n \) via the softmax function [109]:

\[ y_n = \frac{\xi_{p,n} \exp(o_n)}{\sum_k \xi_{p,k} \exp(o_k)} \]  

(5.11)

The weight of a particular class is dependent on the number of samples of that class. If class \( r \) has \( q \) times more samples than those of \( s \), the target is to make one sample of class \( s \) to be as important as \( q \) samples of class \( r \). So, the class weight of \( s \) would be \( q \) times more than the class weight of \( r \).

We optimized the individual models with the Stochastic Gradient Descent (SGD) algorithm. For regularization, we used early-stopping. We split the training dataset into a training set and a cross validation set in 9:1 proportion. Let \( L_{tr}(t) \) and \( L_{va}(t) \) is the average error per example over the training set and validation set respectively, measured after \( t \) epoch. Training was stopped as soon as it reached convergence, i.e., validation error \( L_{va}(t) \) does not improve for \( t \) epoch and \( L_{va}(t) > L_{va}(t-1) \). We used Nesterov momentum optimization with Stochastic Gradient Descent (SGD) algorithm for minimizing the loss of the network. Given an objective function \( f(\theta) \) to be minimized, classic momentum is given by the following pair of equations:

\[ v_t = \mu v_{t-1} - \epsilon \nabla f(\theta_{t-1}) \]  

(5.12a)

\[ \theta_t = \theta_{t-1} + v_t \]  

(5.12b)

where \( v_t \) refers to the velocity, \( \epsilon > 0 \) is the learning rate, \( \mu \in [0,1] \) is the momentum coefficient, and \( \nabla f \theta_t \) is the gradient at \( \theta_t \). On the other hand, Nesterov momentum is given
by:

\[ v_t = \mu v_{t-1} - \epsilon \nabla f(\theta_{t-1} + \mu v_{t-1}) \]  

(5.13a)

\[ \theta_t = \theta_{t-1} + v_t \]  

(5.13b)

The output classification labels of the three individual model are ensembled together using majority voting technique. Each classifier “votes” for a particular class, and the class with the majority votes would be assigned as the label for the input MRI data.

Figure (5.4) Block Diagram of Individual Model \( M_4 \).
5.3 Experiments

We implemented the proposed model using Tensorflow [110], Keras[111] and Python on a Linux X86-64 machine with AMD A8 CPU, 16 GB RAM and NVIDIA GeForce GTX 770. We applied the SGD training with a mini-batch size of 64, a learning rate of 0.01, a weight decay of 0.06 and a momentum factor of 0.9 with Nesterov optimization. We applied early stopping in the SGD training process while there was no improvement (change of less than 0.0001) in validation loss for last six epoch.

To validate the effectiveness of the proposed AD detection and classification model, we developed two baseline deep CNN, Inception-v4 [112] and ResNet [113] and modified their
architecture to classify 3D brain MRI data. Besides, we developed two different models, M_4 and M_5 having similar architecture like M_1, M_2, and M_3 model except for the number of layers in the dense block. M_4, has six (1*1 convolution and 3*3 convolution layers) in the first dense block, twelve (1*1 convolution and 3*3 convolution layers) in the second dense block, forty-eight (1*1 convolution and 3*3 convolution layers) in the third dense block and thirty-two (1*1 convolution and 3*3 convolution layers) in the fourth dense block (Fig. 5.4). The layers in the dense blocks of M_5 has the arrangement 6, 12, 64, 48 as shown in Fig. 5.5. Additionally, we implemented two variants of our proposed model using M_4 and M_5.

- For the first variant, we implemented an ensemble of four deep convolutional neural networks: M_1, M_2, M_3, and M_4. We will refer to this model as E_1.

- For the second variant, we implemented an ensemble system of five deep convolutional neural networks: M_1, M_2, M_3, M_4 and M_5. We will refer to this model as E_2.

Four metrics are used for quantitative evaluation and comparison, including accuracy, positive predictive value (PPV) or precision, sensitivity or recall, and the harmonic mean of precision and sensitivity (f1–score). We denote TP, TN, FP, and FN as True Positive, True Negative, False Positive and False Negative, respectively. The evaluation metrics are defined as:

\[
\text{accuracy} = \frac{(TP+TN)}{(TP+FP+FN+TN)}
\]

\[
\text{precision} = \frac{TP}{(TP+FP)}
\]

\[
\text{recall} = \frac{TP}{(TP+FN)}
\]

\[
\text{f1–score} = \frac{(2TP)}{(2TP+FP+FN)}
\]

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1–Score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Demented</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>73</td>
</tr>
<tr>
<td>Very Mild</td>
<td>0.75</td>
<td>0.50</td>
<td>0.60</td>
<td>6</td>
</tr>
<tr>
<td>Mild</td>
<td>0.62</td>
<td>0.71</td>
<td>0.67</td>
<td>7</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.33</td>
<td>0.50</td>
<td>0.40</td>
<td>2</td>
</tr>
</tbody>
</table>
Table (5.2) Classification Performance of M₂ Model.

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1–Score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Demented</td>
<td>0.88</td>
<td>0.95</td>
<td>0.91</td>
<td>73</td>
</tr>
<tr>
<td>Very Mild</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>6</td>
</tr>
<tr>
<td>Mild</td>
<td>0.25</td>
<td>0.29</td>
<td>0.27</td>
<td>7</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2</td>
</tr>
</tbody>
</table>

5.3.1 Results

We report the classification performance of M₁, M₂, M₃, M₄, and M₅ model in Table 5.1, Table 5.2, Table 5.3, Table 5.4, and Table 5.5, respectively. From the results, we notice that M₁, M₂, and M₃ model are the top performers among all models. So, we choose the ensemble of M₁, M₂, M₃ for our final architecture. Besides, the variants E₁ (M₁+M₂+M₃+M₄), and E₂ (M₁+M₂+M₃+M₄+M₅) demonstrates inferior performance compared to the ensemble of M₁, M₂, M₃ (proposed model) as shown in Fig. 5.6. From Fig. 5.6, we notice that E₁ model has an accuracy of 78% with 68% precision, 78% recall, and 72% f1–score. On the other hand, the E₂ model demonstrates 77% accuracy with 73% precision, 77% recall, and 75% f1–score.

Table (5.3) Classification Performance of M₃ Model.

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1–Score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Demented</td>
<td>0.99</td>
<td>0.96</td>
<td>0.97</td>
<td>73</td>
</tr>
<tr>
<td>Very Mild</td>
<td>0.50</td>
<td>0.33</td>
<td>0.40</td>
<td>6</td>
</tr>
<tr>
<td>Mild</td>
<td>0.45</td>
<td>0.71</td>
<td>0.56</td>
<td>7</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>2</td>
</tr>
</tbody>
</table>

Table (5.4) Classification Performance of M₄ Model.

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1–Score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Demented</td>
<td>0.92</td>
<td>0.67</td>
<td>0.77</td>
<td>73</td>
</tr>
<tr>
<td>Very Mild</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>6</td>
</tr>
<tr>
<td>Mild</td>
<td>0.17</td>
<td>0.60</td>
<td>0.26</td>
<td>7</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2</td>
</tr>
</tbody>
</table>
Table (5.5) Classification Performance of M5 Model.

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1–Score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Demented</td>
<td>0.80</td>
<td>0.94</td>
<td>0.86</td>
<td>73</td>
</tr>
<tr>
<td>Very Mild</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>6</td>
</tr>
<tr>
<td>Mild</td>
<td>0.22</td>
<td>0.14</td>
<td>0.17</td>
<td>7</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2</td>
</tr>
</tbody>
</table>

Table (5.6) Performance of the Proposed Ensembled Model.

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F1–Score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Demented</td>
<td>0.97</td>
<td>1.00</td>
<td>0.99</td>
<td>73</td>
</tr>
<tr>
<td>Very Mild</td>
<td>1.00</td>
<td>0.33</td>
<td>0.50</td>
<td>6</td>
</tr>
<tr>
<td>Mild</td>
<td>0.67</td>
<td>0.86</td>
<td>0.75</td>
<td>7</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5.6 shows the per-class classification performance of our proposed ensembled model on the OASIS dataset [83]. The accuracy of the proposed model is 93.18% with 94% precision, 93% recall and 92% f1–score. The performance comparison of classification results of the proposed ensemble model and the two baseline deep CNN models are presented in Fig. 5.7. Inception-v4 [112] and ResNet [113] have demonstrated outstanding performance for object detection and classification. The reason behind their poor performance for AD detection and classification can be explained by the lack of enough training dataset.

Since these two networks are very deep neural networks, so without a large dataset, training process would not work correctly. On the other hand, the depth of our model is
relatively low, and all the layers are connected to all preceding layers. So, there is a strong gradient flow in times of training that eliminates the ‘Vanishing gradient’ problem. In each training iteration, all the weights of a neural network receive an update proportional to the gradient of the error function concerning the current weight. But in some cases, the gradient will be vanishingly small and consequently prevent the weight from changing its value. It may completely stop the neural network from further training in worst case scenario. Our proposed model does not suffer this ‘Vanishing gradient’ problem, have better feature propagation and provides better classification result even for the small dataset. The performance comparison of classification results of the proposed ensembled model, the baseline deep CNN models and the most recent work, ADNet [81] is presented in Figure 5.8. It can be observed that proposed ensembled model achieves encouraging performance and outperforms the other models.
5.4 Summary

We made an efficient approach to AD diagnosis using brain MRI data analysis. While the majority of the existing research works focuses on binary classification, our model provides significant improvement for multi-class classification. Our proposed network can be very beneficial for early-stage AD diagnosis. Though the proposed model has been tested only on AD dataset, we believe it can be used successfully for other classification problems of medical domain. Moreover, the proposed approach has strong potential to be used for applying CNN into other areas with a limited dataset. In future, we plan to evaluate the proposed model for different AD datasets and other brain disease diagnosis.
Chapter 6

DEEP CONVOLUTIONAL NEURAL NETWORKS FOR AUTOMATED DIAGNOSIS OF ALZHEIMER’S DISEASE AND MILD COGNITIVE IMPAIRMENT

6.1 Introduction

In this chapter, we consider the automated diagnosis of Alzheimer’s Disease (AD) and Mild Cognitive Impairment (MCI) in 3D structural MRI brain scans. We develop an efficient deep convolutional neural network (CNN) based classifier by analyzing 3D brain MRI. The proposed model extracts features from the MRI scans and learns significant information related to Alzheimer’s Disease (AD) and Mild Cognitive Impairment (MCI). We perform motion correction, non-uniform intensity normalization, Talairach transform, intensity normalization, and skull-stripping in the raw MRI scans. After that several 2D slices are generated, and center patch is cropped from the slices before passing them to the CNN classifier. Besides, we demonstrate ways to improve the performance of a CNN classifier for AD and MCI diagnosis. We conduct experiments using Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset for classification of the AD, MCI and CN (normal/healthy controls) to evaluate the proposed model. The proposed model achieves 94.97% accuracy for AD/CN classification and 91.98% accuracy for AD/MCI classification outperforming baseline models and several competing methods from other studies.

AD is the most prevailing type of dementia, and Mild Cognitive Impairment (MCI) is considered as the earlier stage of AD. It is crucial to detect patient at MCI stage before the disease progress further as there is no cure for AD. Earlier diagnosis can help for proper

\footnote{Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf}
treatment and prevent brain tissue damage. The Hippocampus and cerebral cortex of the brain are shrunk, and ventricles are enlarged in the brain of AD patient. Hippocampus reduction causes cell loss and damages the synapses and neuron ends. Structural MRI (sMRI) is helpful for measuring these progressive changes in the brain due to the AD.

6.2 Method

Let \( x = \{x_i, i = 1, ..., N\} \), a set of MRI data with \( x_i \in [0, 1, 2, ..., L - 1]^{h \times w \times l} \), a three dimensional (3D) image with \( L \) gray scale values, \( h \times w \times l \) voxels and \( y \in \{0, 1, 2\} \), one of the stages of AD where 0, 1, 2 refers to normal/healthy control (CN), Mild Cognitive Impairment (MCI), and Alzheimer’s Disease (AD) respectively. We will construct a classifier,

\[
f : X \rightarrow Y; \ x \mapsto y
\]  

that predicts a label \( y \) in response to an input image \( x \) with minimum error rate. The training process of the classifier would be an iterative process to find the set of parameters \( w \), that minimizes the classifier’s loss

\[
L(w, X) = \frac{1}{n} \sum_{i=1}^{n} l(f(x_i, w), \hat{c}_i) \tag{6.2}
\]

where \( x_i \) is \( i^{th} \) image of \( X \), \( f(x_i, w) \) is the classifier function that predicts the class \( c_i \) of \( x_i \) given \( w \), \( \hat{c}_i \) is the ground-truth class for \( i^{th} \) image \( x_i \) and \( l(c_i, \hat{c}_i) \) is the penalty function for predicting \( c_i \) instead of \( \hat{c}_i \). We set \( l \) to the loss of cross-entropy,

\[
l = -\sum_{i} \hat{c}_i \log c_i \tag{6.3}
\]

6.2.1 Data Selection

For our proposed model, we have used 1726 MRI scans (347 AD, 537 CN, 806 MCI) of 479 patients from the Alzheimers Disease Neuroimaging Initiative (ADNI) database
Specifically, we used the ADNI1: Annual 2 Yr 1.5T dataset for our model. The subjects were in the age range 55-92. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Up-to-date information related to ADNI database can be found at www.adni-info.org.

6.2.2 Data Preprocessing

We downloaded the raw Neuroimaging Informatics Technology Initiative (NiFTI) file format MRI scans from the ADNI website (http://adni.loni.usc.edu/). The structural MRI scans were acquired from 1.5T scanners. These MRI scans were already reviewed for quality and Gradient inhomogeneity correction (gradwarp), B1 non-uniformity correction, and N3 processing (to reduce residual intensity non-uniformity) were applied. Since the raw scans are not skull-stripped and have unnecessary information, we perform cortical reconstruction with Freesurfer image analysis suite (http://surfer.nmr.mgh.harvard.edu/). We use the function recon-all -autorecon1 which performs 5 out of 31 transformation processes done by Freesurfer. The five transformation processes are - Motion Correction and Conform, NonUniform intensity normalization (NU), Talairach transform computation, Intensity Normalization 1 and Skull Stripping. After these preprocessing steps we get a skull-stripped MRI scan with dimension 256*256*256. Some slices from sample skull-stripped MRI scan of CN, MCI, and AD patients are shown in Fig. 6.2. We discard several slices at the beginning.

![Figure (6.1) 3D Brain MRI Preprocessing module.](image)
Figure (6.2) Skull-Stripped MRI Slices Presenting Different AD Stages. (a)-(c) CN; (d)-(f) MCI; (g)-(i) AD.

Figure (6.3) Block Diagram of the Proposed Alzheimer’s Disease Diagnosis Framework.

and at the end as they do not have any useful information. After that, we crop a 224*224 center patch from each slice to reduce the background image region outside the brain tissue and perform image normalization.

6.2.3 Data Augmentation

Data augmentation helps to increase the size of the dataset. For our work, we developed an augmentation scheme involving generating multiple slices from each MRI scan. Slices are taken from different image plane: Axial or horizontal plane, Coronal or frontal plane, and Sagittal or median plane. Moreover, we applied Horizontal Flipping to increase the amount of training samples.
Figure (6.4) Deep CNN Architecture Used for the Proposed Alzheimer’s Disease Diagnosis Framework.

### 6.2.4 Network Architecture

Fig. 6.3 shows our proposed model for Alzheimer’s Disease diagnosis. The first stage of the pipeline is the preprocessing module illustrated in Fig. 6.1 and described above. The second stage of the classifier is a deep convolutional neural network as shown in Fig. 6.4. The CNN model is a 2D network and follows a modified architectural pattern of DenseNet-121 [108]. The CNN classifier has several layers performing the convolution, batch normalization, rectified linear unit, and pooling operation. The layers follow a particular connection pattern known as dense connectivity [108]. We keep these layers very narrow (e.g., 12 filters per layer) and connect each layer to every other layer. Similar to [108], we will refer to the layers as dense layer and combination of the layers as dense block. Since all the dense layers are connected to each other, the $i^{th}$ layer receives the feature-maps ($h_0, h_1, h_2, ..., h_{i-1}$), from all previous layers ($0, 1, 2, ..., i-1$). The network has a global feature map set, where each layer adds a small set of feature-maps. Each layer can access the gradients from the loss function.
Table (6.1) Impact of Different Factors on Proposed CNN

<table>
<thead>
<tr>
<th>Methods</th>
<th>AD vs CN</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without Pre-training</td>
<td></td>
<td>75.98</td>
<td>78.45</td>
<td>71.43</td>
</tr>
<tr>
<td>Random slice</td>
<td></td>
<td>75.98</td>
<td>79.51</td>
<td>68.42</td>
</tr>
<tr>
<td>Axial slice</td>
<td></td>
<td>87.15</td>
<td>86.44</td>
<td>88.52</td>
</tr>
<tr>
<td>Sagittal slice</td>
<td></td>
<td>89.94</td>
<td>87.10</td>
<td>96.36</td>
</tr>
<tr>
<td>Coronal slice</td>
<td></td>
<td>94.97</td>
<td>94.33</td>
<td>95.89</td>
</tr>
</tbody>
</table>

Table (6.2) Impact of Different CNN Architecture on Proposed Diagnosis Framework

<table>
<thead>
<tr>
<th>Methods</th>
<th>AD vs CN</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet-18</td>
<td></td>
<td>86.03</td>
<td>85.12</td>
<td>87.93</td>
</tr>
<tr>
<td>ResNet-50</td>
<td></td>
<td>82.68</td>
<td>85.09</td>
<td>78.46</td>
</tr>
<tr>
<td>ResNet-101</td>
<td></td>
<td>83.79</td>
<td>88.18</td>
<td>76.81</td>
</tr>
<tr>
<td>ResNet-152</td>
<td></td>
<td>82.68</td>
<td>79.67</td>
<td>89.28</td>
</tr>
<tr>
<td>DenseNet-169</td>
<td></td>
<td>83.24</td>
<td>84.35</td>
<td>81.25</td>
</tr>
<tr>
<td>DenseNet-201</td>
<td></td>
<td>83.80</td>
<td>88.03</td>
<td>75.81</td>
</tr>
<tr>
<td>Proposed method</td>
<td></td>
<td>94.97</td>
<td>94.33</td>
<td>95.89</td>
</tr>
</tbody>
</table>

and the original input in training time. As a result, the flow of information improves, and gradient flow becomes stronger in the network. Final classification is performed by the softmax layer with three different output classes: CN, MCI, and AD. We optimized the CNN classifier using the Adam algorithm [114].

6.3 Experiments

For our work, we used 80% data from the ADNI1 dataset as training set, and 20% as test dataset. From the training dataset, a random selection of 10% images is used as validation dataset. The experiments were performed using PyTorch framework. Transfer learning [115] was applied to pre-train the CNN classifier using Imagenet database [105]. The parameters used for training process are: learning rate: 0.0001, weight decay: 0.1 after every 7 epochs, and batch size: 16.

To improve the performance of CNN for AD diagnosis from 3D brain MRI, we studied the impact of several factors such as - network pre-training, image pre-processing, choosing
Table (6.3) Comparison with the State-of-the-Art. ‘–’ indicates that result was not reported by the authors.

<table>
<thead>
<tr>
<th>Methods</th>
<th>AD vs CN</th>
<th>AD vs MCI</th>
<th>MCI vs CN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>Sergey et al. [100]</td>
<td>80.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beheshti et al. [116]</td>
<td>93.01</td>
<td>89.13</td>
<td>96.80</td>
</tr>
<tr>
<td>JLRK+DeepESRNET [117]</td>
<td>91.02</td>
<td>92.72</td>
<td>89.94</td>
</tr>
<tr>
<td>MOLK+DeepESRNET [117]</td>
<td>90.28</td>
<td>92.65</td>
<td>89.05</td>
</tr>
<tr>
<td>B. Shi et al. [118]</td>
<td>91.95</td>
<td>89.49</td>
<td>93.82</td>
</tr>
<tr>
<td>Liu et al. [61]</td>
<td>92.0</td>
<td>90.9</td>
<td>93.0</td>
</tr>
<tr>
<td>Aderghal et al. [119]</td>
<td>91.41</td>
<td>89.06</td>
<td>93.75</td>
</tr>
<tr>
<td>Proposed method</td>
<td>94.97</td>
<td>94.33</td>
<td>95.89</td>
</tr>
</tbody>
</table>

random slice from the MRI as network input, and choosing specific slice from three different image plane (axial, sagittal, and coronal).

6.3.1 Results

Initially, we trained the network with raw MRI scans. For this approach, training accuracy was more than 95%, but validation accuracy was around 68% which indicates the network lacks generalization. Pre-processed MRI data helped to solve this issue and improved performance of the CNN. Table 6.1 demonstrates the results of other experiments. From the results, we can see that choosing random slice from the MRI hampers the performance of the CNN classifier even with pre-trained network and pre-processed data. Moreover, the experimental results demonstrate that choosing slices from coronal view have a huge positive impact on the CNN classifier for AD diagnosis. Table 6.2 shows the effect of different CNN architecture on the performance of the proposed AD diagnosis framework. Here, it is evident that the proposed CNN classifier, shown in Fig. 7.2 outperforms the other baseline models. Our model is pre-trained with the Imagenet database [105] and we perform the training with the pre-processed coronal slices. These results also demonstrate that the performance of a CNN classifier vastly depends on the architecture and depth of the network. To validate the effectiveness of our model, we compare it with several state-of-the-art methods. The comparison result is shown in Table 6.3. Following previous approaches, we use accuracy, sensitivity and specificity for performance comparison. The result shows that our model outperforms other competing methods for AD/CN classification, AD/MCI classification and
demonstrates comparable performance for MCI/CN classification.

6.4 Summary

We proposed a novel automated Alzheimer’s Disease diagnosis framework and demonstrated ways to improve the performance of a CNN classifier. The experimental result shows that a pre-trained network with preprocessed slices from coronal view is a reliable technique for MCI and AD diagnosis. The performance of the proposed model shows that it can compete with other state-of-the-art methods for AD diagnosis using 3D brain MRI data.
Chapter 7

GAN-BASED SYNTHETIC BRAIN PET IMAGE GENERATION

7.1 Introduction

In recent days, deep learning technologies have achieved tremendous success in computer vision related tasks with the help of large scale annotated dataset. Obtaining such dataset for medical image analysis is very challenging. Working with the limited dataset and small amount of annotated samples makes it difficult to develop a robust automated disease diagnosis model. We propose a novel approach to generate synthetic medical images using Generative Adversarial Networks (GANs). Our proposed model can create brain PET images for three different stages of Alzheimer’s Disease - Normal Control (CN), Mild Cognitive Impairment (MCI), and Alzheimer’s Disease (AD).

Developing AI assisted automated disease diagnosis systems using medical images often requires a large training dataset with annotated samples, especially for supervised learning methods. Experts with good knowledge of the specific data and task are needed for performing such annotations. So, medical image annotation process is expensive in terms of time, money and effort. It becomes more challenging for precise annotations, such as for identifying different stages of Alzheimer’s Disease. If diagnostic images are intended to be made public, patient consent may be necessary depending on the institutional protocols [120]. So there are very few public medical datasets available online, and they are still limited in size and quality. Collecting medical images for developing automated computer-aided diagnosis system is a complicated and expensive procedure and requires adequate funding, handling

\(^1\)Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf
privacy concern, and collaboration of researchers, physicians, and hospitals. Medical datasets are often imbalanced as pathologic findings are usually rare, and it creates another challenge to train the automated diagnosis system.

![Brain PET images](image)

(a) Sagittal View (b) Coronal View (c) Axial View.

Figure (7.1) Example of Brain PET images (a) Sagittal View (b) Coronal View (c) Axial View.

Data augmentation is one way to overcome the problem of limited dataset. There are several data augmentation techniques, such as translation, rotation, scale, flip, etc. But these techniques are not as useful for medical image analysis as they are for natural image dataset. On the contrary, techniques such as translation and rotation might change the pattern useful for the diagnosis. Besides, these images resemble a great extent to the original ones. So the ML model using these augmented data gain little performance improvements due to the lack of generalization abilities. Another type of data augmentation strategy is synthetic data generation. A synthetic dataset is generated programmatically. Such dataset is highly beneficial for medical image analysis. There is no patient data handling or privacy concerns as the data is produced synthetically. The dataset can contain samples from both positive and negative classes for diagnosis purpose and help build a generalized model.

Generative Adversarial Networks (GANs) [6] can generate synthetic data with good generalization ability. GAN has two different networks - Generator and Discriminator. The model is trained in an adversarial process where the Generator generates fake images, and
the Discriminator learns to discriminate between the real and fake images. There are some excellent research works by the computer vision community to generate synthetic data by using Generative Adversarial Networks [121], [122], [123], [124], [125]. The success of the vision community for synthetic data generation using GAN and the limitation of medical data inspired us to explore methods suitable for medical image synthesis. In this study, we focus on synthetic brain Positron Emission Tomography (PET) image generation for different stages of Alzheimer’s Disease - Normal Control (CN), Mild Cognitive Impairment (MCI), and Alzheimer’s Disease (AD).
7.2 Method

7.2.1 Data Selection

For our proposed model, we have used 411 PET scans (98 AD, 105 CN, 208 MCI) of 479 patients. We collected the data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Specifically, we used ADNI1 baseline dataset for our model. The subjects were in the age range 55-92. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). Up-to-date information related to ADNI database can be found at www.adni-info.org.

7.2.2 Generative Adversarial Networks

Generative Adversarial Networks (GANs) is a deep learning architecture that consisted of two models - a generative model G and a discriminative model D. The generative model
Figure (7.4) Discriminator Architecture of the Proposed Model.

captures the data distribution. The discriminative model estimates the probability that the sample is drawn from the training data rather than the generative model. The two models are simultaneously trained via an adversarial process. The architecture is inspired by game theory and corresponds to a minimax two-player game. The training procedure of $G$ is to maximize the probability of $D$ making a mistake [6].

Let the generator $G(z, \theta_x)$ is a differentiable function represented by a multilayer perceptron with parameters $\theta_g$ that depicts a mapping to the data space. To learn the generator’s distribution $\rho_g$ over the data space $x$, a prior $\rho_z$ is defined on random input noise variables $z$. The discriminator $D(x, \theta_d)$ is also a neural network that gets a sample the real dataset or the generated synthetic dataset produced by $G$ and outputs a single scalar value that the input data comes from the real training dataset. The training process focuses on the task that the discriminator $D$ will maximize the probability of assigning correct labels to the training examples and generated samples from $G$. At the same time, $G$ is trained to generate data samples similar to the real dataset so that $D$ cannot differentiate them from actual data. Similar to game theory, the discriminator $D$ and the generator $G$ play a two-player mini-max game with following value function $V(G, D)$:
\begin{align}
\min_G \max_D V(D, G) &= \mathbb{E}_{x \sim \rho_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim \rho_{\text{data}}(z)} [\log (1 - D(z))] \\
(7.1)
\end{align}

Figure (7.5) Visualization of the Generator Output in the Training Process.

Where \( x \) is the real data and \( z \) is the input random noise. \( \rho_{\text{data}}, \rho_z \) represent the distribution of the real data and the input noise. \( D(x) \) represents the probability that \( x \) came from the real data while \( G(z) \) represents the mapping to synthesize the real data. The generator, \( G \) is a deeper neural network and have more convolutional layers and nonlinearities. The noise vector \( z \) is upsampled while \( G \) learns the weights through backpropagation. At some point, the generator starts producing data that is classified as real by the discriminator.
Deep Convolutional Generative Adversarial Networks (DCGANs)

Deep Convolutional Generative Adversarial Networks (DCGAN) [126] is a major improvement on the first GAN [6]. DCGAN can generate better quality images and have more stability during the training stage. In the synthetic image generation process using the DCGAN, there are two phases: a learning phase and a generation phase. In the training phase, the generator draws samples from an N-dimension normal distribution and works on this random input noise vector by successive upsampling operations, eventually generating an image from it. The discriminator attempts to distinguish between images drawn from the generator and images from the training set [126].

Two important features in DCGAN are BatchNorm ([107] for regulating the extracted feature scale, and LeakyRelu [127] for preventing dead gradients. DCGAN also replace all max pooling with convolutional stride and use transposed convolution for upsampling. It
eliminates fully connected layers and uses Batch normalization. DCGAN uses ReLU in the generator except for the output which uses Tanh and uses LeakyReLU in the discriminator.

7.2.4 Proposed Model

We propose a novel approach to produce synthetic PET images using a Deep Convolutional Generative Adversarial Networks. Following the guidelines to construct the generator and discriminator, described in the paper written by Radford et al. [126], we implemented and trained them on PET scan images using the original discriminator and generator cost functions. Fig. 7.2 shows the proposed synthetic PET image generator model.

**Generator Architecture** The input of the generator is a vector of random 100 numbers drawn from a uniform distribution, and the output is a brain PET image of size 128*128*3. The generator architecture is shown in Fig. 7.3. The network has a fully connected layer and five strided convolutional transpose (known also as ‘deconv’) layers.
strided convolutional transpose layers transform the latent vector into a volume with shape $128 \times 128 \times 3$. Each convolutional-transpose layer is paired with a 2D batch norm layer and a ReLU activation. The strided convolutional transpose layer inserts zeros in between the pixels of the input vector and expands it. The convolution operation is performed over the enlarged vector to create bigger output data. Normalizing responses to have zero mean and unit variance over the entire mini-batch are applied to stabilize the learning process. Fig. 7.5 shows the output of different steps from the generator of the proposed model.

**Discriminator Architecture** The discriminator network consists of a CNN architecture that takes an image of size $128 \times 128 \times 3$ (brain PET image) as input, and outputs one decision: is the brain PET image real or fake? The network consists of five convolution layers with a kernel size of $5 \times 5$ and a fully connected layer. Strided convolutions are applied to each convolutional layer to reduce spatial dimensionality instead of using pooling layers. Batch-normalization and Leaky ReLU activation are applied to each convolutional layer of the network except the output layer. The fully connected output layer has a Sigmoid function to generate the likelihood probability $(0,1)$ score of the input image to be real or fake. The discriminator architecture is shown in Fig. 7.4.

**Training Procedure** We trained the proposed model to synthesize brain PET images for three stages of Alzheimer’s Disease separately. The training process was done iteratively for the generator and the discriminator. We used mini-batches of $m = 64$ brain PET examples for each stage (CN, MCI, and AD) and $m = 64$ noise samples drawn from a uniform distribution between $[-1, 1]$. In the Leaky ReLU, the slope of the leak was set to leak = 0.2. We initialized the weights to a zero-centred normal distribution with a standard deviation of 0.02. Stochastic gradient descent was used in the training process with the Adam optimizer, an adaptive moment estimation that incorporates the first and second moments of the gradients, controlled by parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$ respectively. We applied a learning rate of 0.0001 for 500 epochs.
In the training process, the discriminator is trained to maximize the probability of assigning correct labels to the training examples and the generated samples. At first, the discriminator gets a batch of real samples from the training set. The batch is forward passed through D, and the loss ($\log(D(x)))$ is calculated. The gradients are calculated in a backward pass. Then, a batch of fake samples from the generator is forward passed through D. Similarly, the loss ($\log(1 - D(G(z))))$ is calculated, and the gradients are accumulated with a backward pass. Finally, the gradients from both the all-real and all-fake batches are summed up, and a step of the Discriminators optimizer is done.

The Generator is trained to generate better fake samples by minimizing $\log(1 - D(G(z)))$. The training process maximizes $\log(D(G(z)))$ to minimize the generator’s loss $\log(1 - D(G(z)))$. The output of the generator is passed to the discriminator, and the classification result is collected. The training process repeats unless the generator learns to generate samples labelled as real by the Discriminator.
7.3 Experiments and Results

It is an open issue to develop objective metrics that correlate with perceived quality measurement. For quality evaluation of synthetic images, it should be specific for each application. Following previous state-of-the-art, we performed a quantitative and qualitative assessment of our proposed model. To our best knowledge, no previous works attempted to generate synthetic brain PET images using real PET images. We quantitatively compare the predicted results in terms of Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). PSNR is used to measure the ratio between the maximum possible intensity value and the mean squared error of the synthetic and the real image:

\[
PSNR = 10\log_{10}\left(\frac{(\max(y))^2}{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}\right) \tag{7.2}
\]

where \(n\) is the number of pixels in an image. For our proposed model, the mean PSNR of real and generated images is 32.83.
Structural Similarity Index (SSIM) finds the similarities within pixels of two images; i.e. if the pixels in the two images line up and or have similar pixel density values:

\[
(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}
\]  

where \(x\) is the estimated PET and \(y\) is the ground truth PET, \(\mu_x\) is the average of \(x\), \(\mu_y\) is the average of \(y\), \(\mu_x^2\) is the variance of \(x\), \(\mu_y^2\) is the variance of \(y\), \(\sigma_{xy}\) is the covariance of \(x\) and \(y\). \(C_1 = (k_1L)^2\) and \(C_2 = (k_2L)^2\) are used to stabilize the division with weak denominator where \(L\) is the dynamic range of the pixel-values, \(k_1 = 0.01\) and \(k_2 = 0.03\). For our proposed model, the mean SSIM of real and generated images is 77.48.

We present sample visual results of representative slices from the generated PET data for qualitative comparison. Fig. 7.6, Fig. 7.7, and Fig. 7.8 shows the synthesized PET images from CN, MCI, and AD patients respectively. From the results, we could see that the synthesized brain PET images are quite similar to the real brain PET images. To analyze the similarity between synthetic and real images, we also obtained the 2D-histogram of real and synthetic images. Fig. 7.9 presents the 2D-histogram of a sample real and synthetic image.

There are several limitations to the proposed work. One possible extension could be an increase from 2-D to 3-D input volumes, using 3-D GAN, at the cost of a longer processing time and an increased memory usage. We trained separate GANs for each stage of Alzheimer’s Disease, which increased the training complexity. Future research can focus on the investigation of GAN architectures that generate multi-class samples together.
7.4 Summary

We conclude that synthetic medical image generation is a promising research area and cost-saving approach for developing automated diagnostic technology. Our proposed model can be generalized in other disease diagnosis systems using PET images and can help to supplement the training dataset. The qualitative and quantitative evaluation of the proposed model demonstrates that the synthesised images are close to real brain PET images of different stages of Alzheimer’s Disease. We believe that our proposed model can help to generate labelled images and aid data augmentation for developing robust disease diagnosis systems, and eventually save lives.
Chapter 8

UNDERSTANDING BEHAVIOR OF 3D CONVOLUTIONAL NEURAL NETWORK FOR ALZHEIMER’S DISEASE AND MILD COGNITIVE IMPAIRMENT DIAGNOSIS USING PET DATA

8.1 Introduction

In recent days, Convolutional Neural Networks (CNN) have demonstrated impressive performance in medical image analysis. However, there is a lack of clear understanding of why and how the Convolutional Neural Network performs so well for image analysis task. How CNN analyzes an image and discriminates among samples of different classes are usually considered as non-transparent. As a result, it becomes difficult to apply CNN based approaches in clinical procedure and automated disease diagnosis system. In this paper, we consider this issue and work on visualizing and understanding the decision of Convolutional Neural Network for Alzheimer’s Disease (AD) and Mild Cognitive Impairment (MCI) Diagnosis. We develop a 3D deep convolutional neural network for AD and MCI diagnosis using brain PET scans and propose using five visualizations techniques - Sensitivity Analysis (Backpropagation), Guided Backpropagation, Occlusion, Brain Area Occlusion, and Layer-wise Relevance Propagation (LRP) to understand the decision of the CNN by highlighting the relevant areas in the PET data. We conduct experiments using Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset to train the model and evaluate the visualization techniques.

Alzheimer’s Disease (AD) is a progressive neurodegenerative disease that causes people to lose their memory, mental functions and ability to continue daily activities. AD is the most prevailing type of dementia [119]. It is crucial to detect patient at earlier stage before the disease progress further as there is no cure for AD. Currently, available treatment options for Alzheimer’s Disease focus on slowing the progression of the disease and controlling symp-
toms. So, earlier diagnosis can help for proper treatment and prevent brain tissue damage.

Alzheimer’s Disease is primarily caused by abnormal cell death, mainly in the medial temporal lobe. Such cell death due to the accumulation of the $\beta$-amyloid peptide ($A\beta$) within the brain and the neurofibrillary tangles of hyperphosphorylated tau protein are used to identify Alzheimer’s Disease [128]. These amyloid deposits are caused by the deposition of $A\beta$, a cleavage product of the amyloid precursor protein. The $\beta$-amyloid peptide ($A\beta$) originates from the degenerating mitochondria in dystrophic neurons. Such deposition causes disruption of axons and additional deposition of amyloid [129]. Another important pathophysiological step in the development of Alzheimer’s Disease is the Neurofibrillary tangles caused by the hyperphosphorylation of tau protein [130].

![Figure (8.1) Example of Brain PET Images (a) Normal Control (b) Mild Cognititve Impairment (c) Alzheimer’s Disease.](image)

Positron Emission Tomography is a class of nuclear medicine imaging. It is also known as PET imaging or a PET scan. Nuclear medicine refers to a type of medical imaging methods that utilizes radioactive material in a small amount to diagnose and determine the stage of a disease or treat a disease. Positron Emission Tomography (PET) uses small amounts of radiotracers, a special camera and a computer to evaluate the physiological changes. Positron Emission Tomography (PET) measures the body changes at the cellular level by looking at blood flow, metabolism, neurotransmitters, and radiolabelled drugs. PET may identify early onset of disease before it is evident on other imaging tests. It performs quantitative analyses
Figure (8.2) Skull-Stripped PET Slices.

and finds relative changes over time as the disease process evolves or in response to a specific stimulus [3].

PET imaging tracers can correlate β-amyloid deposition in the brain. The amyloid deposition in the brain can be detected years before the onset of clinical signs of Alzheimer’s Disease. The PET imaging tracers can help in differentiating dementia syndromes at the early stage. PET scans reflect the resting state cerebral metabolic rates of glucose that is an indicator of neuronal activity. Cerebral glucose metabolic alterations have distinct patterns that can be used to identify Alzheimer’s Disease symptoms. Fig. 8.1 shows some brain PET images with different AD stages.

Deep Convolutional Neural Networks have demonstrated immense achievements for image analysis task in recent years, but understanding how they work yet remains a significant challenge. The decision of the Convolutional Neural Networks for image analysis is often considered non-transparent. So it is challenging to build trust for utilizing CNN in developing automated disease diagnosis system. It is essential to diagnose the disease as well as explain the reason for the diagnosis to develop a trustworthy clinical decision support system. To address this issue, understanding the behavior of CNN for classifying medical
images for disease diagnosis is crucial.

Computer Vision community have utilized different visualization methods to understand how Convolutional Neural Network works. Visualizing the learned features of the neurons in different layers of the network is a standard way to understand CNN behavior. A heatmap is generated and projected in the input image highlighting regions that influence the CNN for the classification decision. But there is a lack of such work for medical image analysis using convolutional neural networks. To mitigate this gap, we focus to utilize several visualization methods to understand and explain the behavior of a convolutional neural network for Alzheimer’s Disease and and Mild Cognitive Impairment diagnosis using brain PET data.

![Figure (8.3) Alzheimer’s Disease diagnosis framework using PET data.](image)

8.2 Method

8.2.1 Data Selection

For our proposed model, we have used 1230 PET scans (169 AD, 661 MCI, 400 CN) of 988 patients. We collected the data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Specifically, we used ADNI2 dataset for our model. The subjects were in the age range 55-92. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). Up-to-date information related ADNI database can be found at www.adni-info.org.
Figure (8.4) 3D CNN Classifier for Alzheimer’s Disease and Mild Cognitive Impairment diagnosis using PET data.

8.2.2 Data Preprocessing

PET Florbetapir (formerly AV-45) PET and FDG PET imaging were performed on the patients by ADNI on two separate days (minimum 12hr time lapse). The PET Scans were completed within two weeks before or two weeks after the in-clinic assessments at Baseline and 24 months after Baseline. ADNI 2 subjects had up to 3 florbetapir scans and up to 2 FDG scans, each acquired at 2-year time intervals. We downloaded the raw Neuroimaging Informatics Technology Initiative (NiFTI) file format PET scans from the ADNI website (http://adni.loni.usc.edu/). To ensure the relative alignment among the subjects, we non-linearly registered the PET scans to the 1mm resolution 2009c version of the ICBM152 reference brain using Advanced Normalization Tools (ANTs3). Since the raw scans are not skull-stripped and have unnecessary information, skull stripping and normalization was performed on them. Skull stripping includes removal of non-cerebral tissues like skull, scalp, and dura from brain images and helps reduce computational complexity and time. After this step, all the PET scans resulted in volumes of 193 * 229 * 193. Some slices from the skull-stripped PET scan are shown in Fig. 8.2.
8.2.3 CNN Architecture

Let $x = \{x_i, i = 1, ..., N\}$, a set of brain PET data with $x_i \in [0, 1, 2, ..., L - 1]^{h*w*l}$, a three Dimensional (3D) image with $L$ gray scale values, $h*w*l$ voxels and $y \in \{0, 1, 2\}$, one of the stages of AD where 0, 1, and 2 refers to normal/healthy control (CN), Mild Cognitive Impairment (MCI) and Alzheimer’s Disease (AD) respectively. We constructed a classifier,

$$f : X \rightarrow Y; \ x \mapsto y \tag{8.1}$$

that predicts a label $y$ in response to an input image $x$ with minimum error rate. The training process of the classifier is an iterative process to find the set of parameters $w$, that
minimizes the classifier’s loss

\[
L(w, X) = \frac{1}{n} \sum_{i=1}^{n} l(f(x_i, w), \hat{c}_i)
\]  \hfill (8.2)

where \( x_i \) is \( i^{th} \) image of \( X \), \( f(x_i, w) \) is the classifier function that predicts the class \( c_i \) of \( x_i \) given \( w \), \( \hat{c}_i \) is the ground-truth class for \( i^{th} \) image \( x_i \) and \( l(c_i, \hat{c}_i) \) is the penalty function for predicting \( c_i \) instead of \( \hat{c}_i \). We set \( l \) to the loss of cross-entropy,

\[
l = -\sum_i \hat{c}_i \log c_i
\]  \hfill (8.3)

Fig. 8.3 represents the block diagram of the Alzheimer’s Disease diagnosis model using PET data. We developed a 3D Convolutional Neural Network inspired by the architecture [131]. The CNN architecture is depicted in Fig. 8.4. The CNN classifier has several layers performing the convolution, batch normalization, rectified linear unit, and pooling operation. There are four convolutional layers and three fully connected layers. Final classification is performed by the softmax layer with three different output classes: CN, MCI, and AD. We optimized the CNN classifier using the Adam algorithm [114]. Though the focus of current work is CNN visualization and not the classification performance, the model achieves a classification performance of 88.76%.

8.2.4 Visualization Methods

**Sensitivity Analysis (Backpropagation)** Simonyan et al. [95] proposed an Image-Specific Class Saliency Visualisation method that ranks the influence of an image, \( x \), according to its impact on a specific class score. The approach finds the relationship between an output class score and the input \( x \) as the prediction of a specific class, \( y \). Given an image \( I_0 \), a class \( c \), and a classification ConvNet with the class score function \( S_c(I) \), the pixels of \( I_0 \) are ranked based on their influence on the score \( S_c(I_0) \). For CNN, the class score \( S_c(I) \) is a highly non-linear function of \( I \). The highly non-linear function of \( y = S_c(I) \) is approximated
to be:

\[ S_c(I) \approx w^T I + b \]  

(8.4)

where \( b \) is the bias of the model and \( w \) is the derivative of \( S_c \) with respect to the image \( I \) at the point (image) \( I_0 \):

\[ w = \frac{\delta S_c}{\delta I}_{|I_0} \]  

(8.5)

Figure (8.6) Relevance Heatmaps for Sensitivity Analysis (Backpropagation) Method Averaged Over MCI Patients.

**Guided Backpropagation (GB)** Guided backpropagation is a slightly modified version of saliency extraction method proposed by Simonyan et al. [95]. Springenberg et al. [96] proposed Guided backpropagation (GB) algorithm where the heatmap is generated using the absolute values of the gradient of the output concerning the input nodes. The negative gradients are set to zero at the rectification layers of the network to focus on what image features the neuron detects. The GB method updates the gradient \( \delta \), the derivative with
respect to the input (or activations for gradients of the intermediate layers) $x$, of ReLUs during the backpropagation. A ReLU unit, $y = \max(0, x)$ backpropagates the gradient of the layer next to the ReLU. The positive portion of the input $\delta_{i-1} = \delta_i$, is backpropagated where $x > 0$, $\delta_i > 0$, $\delta_i$ is the gradient of the layer succeeding the ReLU, and $\delta_{i-1}$ is the gradient after back propagating through ReLU.

**Occlusion Sensitivity** Zeiler et al. [94] proposed an Occlusion Sensitivity method to understand the function of intermediate feature layers and the operation of a CNN classifier. To interpret the feature activity in the intermediate layers of a CNN, they proposed to map these activities back to the input pixel space. Thus they identified the input pattern that initially caused a given activation in the feature maps. They also systematically occluded different portions of the input image with a grey square and monitored the output of the classifier. If the probability for the target class decreased compared to the original image, this image region is considered to be relevant for that target class. To generate a relevance heatmap, the patch is moved across the image, and the difference is calculated between unoccluded and occluded probability.

**Brain Area Occlusion** This approach is inspired by a segmentation-based visualization approach proposed by Yang et al. [97]. We occlude an entire brain area based on the Automated Anatomical Labeling atlas (AAL, http://www.gin.cnrs.fr/en/tools/aal-aal2/) following [132] and calculate the difference between unoccluded and occluded probability.

**Layer-wise Relevance Propagation (LRP)** Bach et al. [99] proposed Layer-wise Relevance Propagation (LRP) algorithm to understand the classification decisions of the model by pixel-wise decomposition. They attempted to visualize the contributions of single pixels to predictions for kernel-based classifiers over Bag of Words features and for multi-layered neural networks. Heatmaps can be produced from the pixel visualization of LRP method and provided to human expert for validation and analysis. The approach focuses on
finding relevance to individual input nodes and trace back contributions to the final output node in a layer by layer fashion. Each of the nodes in layer $l$, that contributes to the activation of a node $j$ in layer $l + 1$ has a part in the relevance $R_{i+1}^j$. The total relevance or the activation strength of an output node for a particular class is computed as:

$$\sum_i R_{i,l+1}^{i\rightarrow j} = R_{l+1}^j$$  \hspace{1cm} (8.6)

8.3 Experiments and Results

For our work, we used 80% data from the dataset as the training set, and 20% as test dataset. From the training dataset, a random selection of 10% images is used as validation dataset. The training-test and training-validation split were done at the subject/patient level. The experiments were performed using PyTorch framework. The parameters used for the training process are learning rate: 0.0001, weight decay: 0.1 after every seven epochs, and batch size: 16.

8.3.1 Classification Result

The network achieves a comparable classification accuracy of 88.76% for CN/AD classification and 64.57% for CN/MCI classification. Please note that the focus of the current study is CNN visualization and not classification performance. In our experiment, we also developed a 2D-CNN model using axial, coronal, and sagittal slices from PET data that achieved 71.45% classification accuracy for CN/AD classification. The vast difference in the classification result suggests that 3D-CNN networks have better capability to learn features from three-dimensional PET image data. Though to validate such findings, further experiments are necessary.

8.3.2 Visualization

We generated relevance relevance heatmaps for all visualizations methods, averaged over CN, MCI, and AD PET images in the test dataset. Fig. 8.5, Fig. 8.6, and Fig. 8.7 repre-
Figure (8.7) Relevance Heatmaps for Sensitivity Analysis (Backpropagation) Method Averaged Over AD Patients.

...sents the relevance heatmaps generated for Sensitivity Analysis (Backpropagation) Method for CN, MCI, and AD patients respectively. Fig. 8.8, Fig. 8.9, and Fig. 8.10 represents the relevance heatmaps generated for Guided Backpropagation Method for CN, MCI, and AD patients respectively. Fig. 8.11, Fig. 8.12, and Fig. 8.13 represents the relevance heatmaps generated for Occlusion Method for CN, MCI, and AD patients respectively. Fig. 8.14, Fig. 8.15, and Fig. 8.16 represents the relevance heatmaps generated for Brain Area Occlusion Method for CN, MCI, and AD patients respectively. Fig. 8.17, Fig. 8.18, and Fig. 8.19 represents the relevance heatmaps generated for LRP Method for CN, MCI, and AD patients respectively. Fig. 8.20 presents the visual comparison of these five methods. The red areas/dots indicate that regions were important for the decision making of the 3D-CNN model.

From the result, we can see that all the visualization focuses mostly on similar brain regions and the generated heatmaps for CN, MCI, and MCI patients are similar. It indicates...
that the network focuses on the same regions to identify the stage of Alzheimer’s Disease. There are some differences, such as the heatmaps generated for the gradient-based methods are distributed. The heatmaps highlight the areas that the CNN network is most susceptible. For the LRP method, the heatmap shows the average relevance of each voxel for contributing to the Alzheimer’s Disease diagnosis score. The heatmaps generated by the occlusion based methods are more focused on the specific regions and cannot administer with large areas of distributed relevance. The reason behind the issue is the occlusion path were not able to cover those areas (for example, the cortex) completely. Brain area occlusion presents very high relevance for the temporal lobe. Since in this method, only one area is covered at a time, that can cause such high importance for one region and minimal relevance for other areas.

Figure (8.8) Relevance Heatmaps for Guided Backpropagation Method Averaged Over CN Patients.
8.3.3 Impact of Brain Region

To evaluate the visualization result, we consider the literature available from the medical domain [133], [134], as there is no ground truth for validating the generated heatmaps. The top five most relevant regions for each visualization method is shown in Table. 8.1. The relevance in each area following the AAL atlas was summed together to identify the most relevant regions for CN, MCI, and AD PET scans. From the top five relevant brain area for each visualization method, we can see that the 3D-CNN focuses on the temporal lobe area, including hippocampus for CN/MCI/AD classification. So, our findings are similar to those in medical literature [133], [134].
8.4 Summary

We developed a 3D Convolutional Neural Network for Alzheimer’s Disease and Mild Cognitive Impairment classification and applied five visualization methods to understand the network’s behavior to classify sample brain PET data. The 3D-CNN focuses on the temporal lobe area, including hippocampus for CN/MCI/AD classification. Our findings are similar to those in medical literature, and these biomarkers were identified directly from the network’s decision without having any ground truth. Such findings are crucial for building trust in clinical practitioners for adapting automated disease diagnosis systems. Our proposed approach can be extended for understanding CNN behavior for other disease diagnosis systems too. In future, we plan to analyze the generated heatmaps for the misclassified samples and identify the regions and patterns that can cause false positive and false negative results.
Figure (8.11) Relevance Heatmaps for Occlusion Method Averaged Over CN Patients.

Figure (8.12) Relevance Heatmaps for Occlusion Method Averaged Over MCI Patients.
Figure (8.13) Relevance Heatmaps for Occlusion Method Averaged Over AD Patients.

Table (8.1) Brain Area Relevance for CN, MCI, and AD Patient for Each Visualization Method

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity Analysis (Backpropagation)</th>
<th>Guided Backpropagation</th>
<th>Occlusion Sensitivity</th>
<th>Brain Area Occlusion</th>
<th>Layer-wise Relevance Propagation (LRP)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CN</strong></td>
<td>Frontoal_Mid (5.78%)</td>
<td>Frontal_Mid (6.31%)</td>
<td>Precuneus (7.96%)</td>
<td>Precuneus (14.90%)</td>
<td>Cerebral White Matter (21.7%)</td>
</tr>
<tr>
<td></td>
<td>Precuneus (5.01%)</td>
<td>Precuneus (4.87%)</td>
<td>Postcentral (7.09%)</td>
<td>Frontal_Mid (9.14%)</td>
<td>Cerebellum (11.74%)</td>
</tr>
<tr>
<td></td>
<td>Postcentral (4.75%)</td>
<td>Postcentral (4.70%)</td>
<td>Precuneus (5.11%)</td>
<td>Postcentral (7.11%)</td>
<td>Frontal_Mid (5.83%)</td>
</tr>
<tr>
<td></td>
<td>Temporal_Mid (4.49%)</td>
<td>Precuneus (4.23%)</td>
<td>Temporal_Mid (5.03%)</td>
<td>Temporal_Inf (5.62%)</td>
<td>Precuneus (3.72%)</td>
</tr>
<tr>
<td></td>
<td>Precentral (4.05%)</td>
<td>Temporal_Mid (3.74%)</td>
<td>Parietal_Sup (4.39%)</td>
<td>Frontal_Mid (5.25%)</td>
<td>Frontoal_Mid (3.32%)</td>
</tr>
<tr>
<td><strong>MCI</strong></td>
<td>Frontoal_Mid (5.23%)</td>
<td>Frontal_Mid (5.61%)</td>
<td>Temporal_Mid (7.69%)</td>
<td>Temporal_Mid (15.94%)</td>
<td>Cerebral White Matter (42.69%)</td>
</tr>
<tr>
<td></td>
<td>Temporal_Mid (4.61%)</td>
<td>Occipital_Mid (4.60%)</td>
<td>Postcentral (7.09%)</td>
<td>Postcentral (7.89%)</td>
<td>Cerebellum (7.82%)</td>
</tr>
<tr>
<td></td>
<td>Precuneus (4.48%)</td>
<td>Precuneus (4.48%)</td>
<td>Precuneus (5.91%)</td>
<td>Occipital_Mid (7.89%)</td>
<td>Frontal_Mid (3.43%)</td>
</tr>
<tr>
<td></td>
<td>Postcentral (4.36%)</td>
<td>Postcentral (5.41%)</td>
<td>Precuneus (5.41%)</td>
<td>Precuneus (7.53%)</td>
<td>Angular Gyr. (3.10%)</td>
</tr>
<tr>
<td></td>
<td>Occipital_Mid (3.92%)</td>
<td>Precuneal (4.66%)</td>
<td>Precuneal (5.97%)</td>
<td>Precentral (5.97%)</td>
<td>Lateral Ventricles (2.99%)</td>
</tr>
<tr>
<td><strong>AD</strong></td>
<td>Frontal_Mid (5.59%)</td>
<td>Frontal_Mid (5.92%)</td>
<td>Postcentral (8.77%)</td>
<td>Precuneus (14.11%)</td>
<td>Cerebral White Matter (32.60%)</td>
</tr>
<tr>
<td></td>
<td>Precuneus (4.87%)</td>
<td>Postcentral (4.72%)</td>
<td>Precuneus (8.14%)</td>
<td>Temporal_Mid (13.27%)</td>
<td>Cerebellum (9.78%)</td>
</tr>
<tr>
<td></td>
<td>Postcentral (4.82%)</td>
<td>Precuneus (4.55%)</td>
<td>Temporal_Inf (9.51%)</td>
<td>Temporal_Mid (5.51%)</td>
<td>Frontal_Mid (3.26%)</td>
</tr>
<tr>
<td></td>
<td>Temporal_Mid (4.57%)</td>
<td>Precuneal (3.93%)</td>
<td>Parietal_Sup (5.01%)</td>
<td>Angular (4.79%)</td>
<td>Lateral Ventricles (3.02%)</td>
</tr>
<tr>
<td></td>
<td>Precentral (4.32%)</td>
<td>Temporal_Mid (4.82%)</td>
<td>Temporal_Mid (5.01%)</td>
<td>Frontal_Mid (5.51%)</td>
<td>Brain Stem (2.95%)</td>
</tr>
</tbody>
</table>
Figure (8.14) Relevance Heatmaps for Brain Area Occlusion Method Averaged Over CN Patients.
Figure (8.15) Relevance Heatmaps for Brain Area Occlusion Method Averaged Over MCI Patients.
Figure (8.16) Relevance Heatmaps for Brain Area Occlusion Method Averaged Over AD Patients.
Figure (8.17) Relevance Heatmaps for Layer-wise Relevance Propagation (LRP) Method Averaged Over CN Patients.
Figure (8.18) Relevance Heatmaps for Layer-wise Relevance Propagation (LRP) Method Averaged Over MCI Patients.
Figure (8.19) Relevance Heatmaps for Layer-wise Relevance Propagation (LRP) Method Averaged Over AD Patients.
Figure (8.20) Visualization Comparison of the Relevance Heatmaps. (a) Sensitivity Analysis (Backpropagation). (b) Guided Backpropagation. (c) Occlusion. (d) Brain Area Occlusion. (e) Layer-wise Relevance Propagation (LRP).
Chapter 9

FUTURE WORK

In this section, we are presenting more future problems, challenges, and pointing out the future research directions.

9.1 RESEARCH DIRECTIONS

In this dissertation, we have addressed several challenges related to medical image analysis with deep learning technologies for developing better disease diagnosis model. Most of our work used 2D-CNN utilizing slices from the axial, coronal, sagittal view of 3D MRI data. While working for CNN visualization, we developed two 3D-CNN models for MRI and PET data analysis, and they showed promises for better diagnosis result. In future, we plan to utilize the 3D-CNN model for analyzing medical images for other neurological disorders.

Developing a multimodal system for diagnosing disease using data from different modalities, such as PET and MRI, could improve the diagnosis process. In future, we plan to develop such systems that can utilize data from different modalities and offer better diagnosis result.

We have generated synthetic data for different stages of Alzheimer’s disease using Generative Adversarial networks. Such data will help a lot to increase the volume of the training set for training a model to identify the progression of the disease. Another exciting research direction would be generating data samples from one modality to another, such as - MRI to PET or vice versa. For synthetic medical data generation, one possible extension could be an increase from 2D to 3D input volumes, using 3D GAN. We trained separate GANs for each stage of Alzheimer’s Disease, which increased the training complexity. Future research can focus on the investigation of GAN architectures that generate multi-class samples together. Additionally, we plan to apply our model for generating synthetic data for other diseases,
especially neurological disorders.

In this dissertation, we have generated heatmaps for understanding CNN behavior for Alzheimer’s Disease Diagnosis. Generating ground truth for such relevance heatmaps to identify Alzheimer’s Disease progression could be another exciting research direction. Furthermore, to analyze the generated heatmaps for the misclassified samples and identify the regions and patterns that can cause false positive and false negative results can help developing better disease diagnosis model.
Chapter 10

CONCLUSIONS

This dissertation has presented our examination of learning deep feature representations for medical image analysis to develop better disease diagnosis system. We have presented approaches for diagnosis of Alzheimer’s disease from brain MRI and PET data and presented strong results showing the effectiveness of these methods.

First, we made an efficient approach to AD diagnosis using brain MRI data analysis. Our proposed network can be very beneficial for early-stage Alzheimer’s Disease diagnosis. Though the proposed model has been tested only on Alzheimer’s Disease dataset, we believe it can be used successfully for other classification problems of medical domain. Moreover, the proposed approach has strong potential to be used for applying CNN into other areas with a limited dataset.

Second, we demonstrated ways to improve the performance of a CNN classifier. The experimental result shows that a pre-trained network with preprocessed slices from coronal view is a reliable technique for MCI and AD diagnosis. The performance of the proposed model shows that it can compete with other state-of-the-art methods for AD diagnosis using 3D brain MRI data.

Third, we proposed approaches for synthetic medical image generation to solve the limited dataset problem for developing automated diagnosis model. Our proposed model can be generalized in other disease diagnosis systems using PET images and can help to supplement the training dataset. The qualitative and quantitative evaluation of the proposed model demonstrates that the synthesised images are close to real brain PET images of different stages of Alzheimer’s Disease. We believe that our proposed model can help to generate labelled images and aid data augmentation for developing robust disease diagnosis systems.
Fourth, we developed a 3D Convolutional Neural Network for Alzheimer’s Disease and Mild Cognitive Impairment classification and applied five visualization methods to understand the network’s behavior to classify sample brain PET data. The 3D-CNN focuses on the temporal lobe area, including hippocampus for CN/MCI/AD classification. Our findings are similar to those in medical literature, and these biomarkers were identified directly from the network’s decision without having any ground truth. Such findings are crucial for building trust in clinical practitioners for adapting automated disease diagnosis systems. Our proposed approach can be extended for understanding CNN behavior for other disease diagnosis systems too.

To conclude, this dissertation proposed approaches that have the potential to improve the development of automated disease diagnosis system, especially Alzheimer’s Disease and Mild Cognitive Impairment. We believe our methods can be utilized for medical image analysis of other diseases to perform accurate and early diagnosis, and eventually save lives.
REFERENCES


[118] B. Shi, Y. Chen, P. Zhang, C. D. Smith, J. Liu, A. D. N. Initiative et al., “Nonlin-


