Computer-Aided Diagnosis for Early Identification of Multi-Type Dementia using Deep Neural Networks

by

Elham Alkabawi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2017

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

With millions of people suffering from dementia worldwide, the global prevalence of this condition has a significant impact on the global economy. As well, its prevalence has a negative impact on patients' lives and their caregivers' physical and emotional states. Dementia can be developed as a result of some risk factors as well as it has many forms whose signs are sometimes similar. While there is currently no cure for dementia, effective early diagnosis is essential in managing it. Early diagnosis helps people in finding suitable therapies that reduce or even prevent further deterioration of cognitive abilities, and in taking control of their conditions and planning for the future. Furthermore, it also facilitates the research efforts to understand causes and signs of dementia. Early diagnosis is based on the classification of features extracted from three-dimensional brain images. The features have to accurately capture main dementia-related anatomical variations of brain structures, such as hippocampus size, gray and white matter tissues' volumes, and brain volume.

In recent years, numerous researchers have been seeking the development of new or improved Computer-Aided Diagnosis (CAD) technologies to accurately detect dementia. The CAD approaches aim to assist radiologists in increasing the accuracy of the diagnosis and reducing false positives. However, there is a number of limitations and open issues in the state-of-the-art, that need to be addressed. These limitations include that literature to date has focused on differentiating multi-stage of Alzheimer's disease severity ignoring other dementia types that can be as devastating or even more. Furthermore, the high dimensionality of neuroimages, as well as the complexity of dementia biomarkers, can hinder classification performance. Moreover, the augmentation of neuroimaging analysis with contextual information has received limited attention to-date due to the discrepancies and irregularities of the various forms of data. This work focuses on addressing the need for differentiating between multiple types of dementia in early stages.

The objective of this thesis is to automatically discriminate normal controls from patients with various types of dementia in early phases of the disease. This thesis proposes a novel CAD approach, integrating a stacked sparse auto-encoder (SSAE) with a two-dimensional convolutional neural network (CNN) for early identification of multiple types of dementia based on using the discriminant features, extracted from neuroimages, incorporated with the context information. By applying SSAE to intensities extracted from magnetic resonance (MR) neuroimages, SSAE can reduce the high dimensionality of neuroimages and learn changes, exploiting important discrimination features for classification. This research work also proposes to integrate features extracted from MR neuroimages with patients' contextual information through fusing multi-classifier to enhance the early prediction of various types of dementia.

The effectiveness of the proposed method is evaluated on OASIS dataset using five different relevant performance metrics, including accuracy, f1-score, sensitivity, specificity, and precision-recall curve. Across a cohort of 4000 MR neuroimages (176×176) as well as the contextual information, and clinical diagnosis of patients serving as the ground truth, the proposed CAD approach was shown to have an improved F-measure of 93% and an average area under Precision-Recall curve of 94%. The proposed method provides a significant improvement in classification output, resulted in high and reproducible accuracy rates of 95% with a sensitivity of 93%, and a specificity of 88%.

Acknowledgement

My sincere gratitude and thanks go to Almighty God for giving me the knowledge, opportunity, and the strength to accomplish this work.

Along the way, several people deserve sincere recognition. I would like first to express my sincere gratitude to my advisors, Prof. Otman Basir and Dr. Allaa Hilal, for their valuable advice and guidance and for their generous help. Their constant support was the secret of achieving all this work. I express my deepest appreciation for granting me the opportunity to be a graduate student at the University of Waterloo. Their valuable guidance, keen interest, and continuous encouragement helped me during my research.

I would like to extend my appreciation and gratitude to my thesis committe members, Dr. Hamid Tizoosh, and Dr. Mark Crowley for sparing their time for reviewing this thesis and for their insightful comments. I would like also to acknowledge the Saudi Ministry of Higher Education and Saudi Cultural Bureau for supporting me financially and logistically to finish this research work. I would like to thank all UW staff members who helped directly during my study.

I would like to thank all my family members. My mom, Zahra Marzouq, and my brothers and sisters, Ahmed, Fatimah, Ali, and Zainab, who have been the greatest supporters in my life. I am grateful to them for their unconditional love and support.

Dedication

This thesis is dedicated to Allah. I praise him for his guidance and blessings.

To the memory of my beloved father, for his guidance to me to become a successful woman in future and teaching me to be self-reliant. May his soul rest in peace.

To my lovely mother, for her prayers for me and her always presence besides me.

To my dear brothers and sisters, for their unconditional love and support.

To my close friends, for their unconditional motivation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Dementia is a brain disorder characterized by a chronic decline in mental ability due to loss of, or damage to, neurons in the brain. The term dementia is widely used for brain disorders that result in a set of symptoms that disturb normal brain functions. These include thinking, intellectual abilities, memory recollection, problem-solving, and use of language. Dementia is severe enough to affect the patients' daily activities. In recent years, the incidence of dementia has increased at an incredible rate. Every year, 7.7 million new cases are diagnosed as having dementia worldwide. In [89], the article reported that 47.5 million people have dementia worldwide, with just over half (58 %) living in low- and middle-income countries. The global prevalence of dementia has had a significant impact on patients' lives, their caregivers' physical and emotional states, and the global economy. In 2015, total health care cost of people with dementia amounted more than 1 percent of the global gross domestic product or US$ 818 billion. Early diagnosis of dementia is essential to help people to take control of their condition, plan for the future, and get proper treatment to delay the progression of the disease into advanced stages, as well it facilitates the research efforts to understand the causes of dementia [39]. Dementia has many types who often present themselves with very similar symptoms. These types can be split into two groups based on which part of the brain is affected. Cortical dementias happen because of problems in the cerebral cortex, whereas subcortical dementias happen because of problems in the parts of the brain beneath the cortex. When studying medical neuroimages with various dementia types, many differential clinical and biochemical markers can be observed with respect to the cerebral cortex, hippocampus, ventricles, and gray and white matter tissues.

Accordingly, numerous researchers have been working towards the development of new or improved CAD technologies to accurately detect dementia. There is a number of limitations in the state-of-the-art, and there are many gaps that need to be addressed. This includes the state-of-the-art studies focused on differentiating various stages of severity of Alzheimer' s disease (AD), whereas there remains a gap in the early diagnosis of other types of dementia that can be as devastating or even more. Moreover, the high dimensionality of neuroimages, as well as the extreme complexity of dementia biomarkers, can hinder classification performance. Furthermore, the combination of medical images and the contextual data is challenging. However, this work focuses on early identification of various types of dementia using medical neuroimages.

This thesis proposes a novel deep-learning-based CAD approach with an ensemble of classifiers to differentiate among multiple types of dementia in the early stages. This research work proposes using dimensionality reduction techniques to reduce images' high dimensionality and extract key features and then accurately classify different forms of dementia with lower computational costs and lower storage space. Thus, the dimensionality reduction helps in making the learning model more robust and faster. Furthermore, this study aims to enhance classification performance by fusing the classification outputs of two individual classifiers.

1.2 Objectives and Hypothesis

This study aims to help radiologists in obtaining early accurate predictions of various types of dementia. Section 1.2.1 summarizes the research objectives. These objectives then assisted in formulating hypotheses, which will be tested in this research. These hypotheses are presented in Section 1.2.2.

1.2.1 Research Objectives

This thesis aims to accurately classify several types of dementia in early stages. To accomplish this goal, this research work seeks to provide a fast and a robust CAD model for the classification task by reducing the high dimensionality of neuroimages while extracting discriminant features . Thus, dimensionality reduction allows for multiple types of dementia to be classified with lower computational costs and storage space. Furthermore, the research work aims to enhance the accuracy of the model for predicting multiple types of dementia by integrating features extracted from neuroimages and contextual data through the fusion of multi-classifiers.

1.2.2 Research Hypotheses

This research work seeks to prove some hypotheses using some experiments. These hypotheses are summarized as follows:

- Predicting multiple types of dementia with deep learning of dementia biomarkers could be achieved with very high classification accuracy.

- Reducing high dimensionality of neuroimages into high-level feature would yield higher classification performance and accurate identification of various types of dementia biomarkers.

- Incorporating medical neuroimages with the clinical and physiological data, utilized by many classifiers, would enhance the accuracy of the diagnosis and outcome early prediction of several types of dementia.

1.3 Contributions

The proposed research offers a deep-learning-based CAD approach using the Convolutional Neural Network (CNN) and logistic regression for distinguishing multiple types of dementia. The most prominent contributions of this research paper include:

- **The Diagnosis of Various Types of Dementia using a convolutional neural network:**

  State-of-the-art research has focused on differentiating multiple stages of Alzheimer's disease, ignoring other dementia types, i.e., vascular dementia and frontotemporal lobe degeneration, that can be as devastating or even more. In this research, a novel convolutional neural network (CNN)-based approach is proposed for the early diagnosis of multiple types of dementia. The CNN takes images' pixel intensities to learn high-level features, representing dementia characteristics. Then, the learned features are classified by feeding them into the output layer of the CNN that recognizes the type of dementia. Our results perform better than results of comparing state-of-the-art approaches.

- **Feature extraction and dimensionality reduction using a Stacked Sparse Autoencoder (SSAE):**

  Neuroimages may suffer from the curse of high dimensionality that makes it difficult to visualize and capture features required for classification during the analysis phase. In fact, high dimensionality of images has a negative impact on the performance of Machine Learning (ML) classifiers during training on these images, so that overfitting increases due to increased data complexity. To address this concern, the SSAE is proposed for reducing high dimensionality of data and extracting relevant features to provide a rapid and robust model for predicting dementia in its early stages, and then improving accuracy and efficiency of the model for early identification of five types of dementia.

- **Incorporation of contextual data through multiple classifiers fusion:**

  Different types of dementia share the same set of pathological characteristics. This is further compounded by the need for early diagnosis in which medical brain scans can show limited biomarkers. Thus, neuroimages may not be sufficient for accurate identification of dementia. This work proposes a novel context-aware CAD for early detection of multiple types of dementia by augmenting the medical imaging data with contextual patients' data through the decision fusion of multiple classifiers to enhance the accuracy of identification of the various types of dementia.

1.4 Thesis Outline

The structure of this thesis can be broken down into six chapters.

Chapter 2 provides a general overview of dementia by explaining the statistics of its prevalence, reviewing the risk factors that increase the chances of developing dementia, and discussing the most common types of dementia. It also covers an overview of all medical tests used in diagnosing dementia as well as a brief study of the most popular performance measures used to evaluate the CAD methodologies in diagnosing dementia. Furthermore, the most common features that have been used in state-of-the-art CAD methodologies to enhance early detection of dementia have been highlighted. This chapter also provides a comprehensive review of the CAD methodologies that have been implemented for early detection of dementia.

Chapter 3 proposes a novel deep-learning-based CAD approach for classification of multiple types of dementia. The proposed approach utilizes the CNN to extract features, which are then used for classifying MR brain images into one of the various types of dementia. Then, the classification accuracy is computed for evaluating the proposed approach's performance and comparing it to the performance results of some state-of-the-art techniques in the detection of dementia.

Chapter 4 advances a deep-learning-based algorithm previously proposed using an auto-encoder that seeks to learn a subset of relevant features to facilitate learning and improve classification performance. Then, the CNN is trained with high-level features of MRI images for classification. The performance of the proposed learning approach is verified by using various classifiers with both the SAE-based low dimensional dataset and with PCA-driven representation for comparison.

Chapter 5 advances the learning approach provided in Chapter 4 to enhance the performance of early detection of multiple types of dementia. It offers a novel CAD system for early detection of multiple types of dementia that initially pre-trains SSAEs to extract features, denoting differentiable structural variations in neuroimages. Subsequently, the extracted features are classified using CNN. As well, the extracted features, as well as context, are classified by logistic regression into categories of dementia types. Probabilities, outputted by two classifiers are fused by linear pooling decision fusion to obtain the final decision probability.

Finally, Chapter 6 summarizes the study' s results presented in Chapter 3, Chapter 4, and Chapter 5 as well as presents possibilities for future research work.

Chapter 2

Background

2.1 Introduction

Dementia is a brain disorder that is characterized by a chronic decline in mental ability due to loss of, or damage to, neurons in the brain. The term dementia is a widespread term for brain disorders that result in a set of symptoms that disturb normal brain functions, such as thinking, intellectual abilities, memory recollection, problem-solving, and use of language, serious enough to affect their daily activities [23]. In recent years, the number of patients who have Alzheimer's disease (AD) and other dementias indicates the prevalence of such cognitive impairment. Early diagnosis of dementia is intended to help people to get proper treatments, prevent or slow down cognitive ability deterioration, as well as seek the appropriate support and plan for the future [39]. Accordingly, numerous researchers have been seeking the development of new or improved technologies to accurately detect dementia. The most prominent diagnostic technique that can capture early symptoms of dementia is through revealing the internal structure and the function of the brain. This paper provides a novel comprehensive study of the state-of-the-art CAD methodologies that enable early detection of dementia.

This chapter provides a novel comprehensive study of the state-of-the-art CAD methodologies that enable early detection of dementia. Section 2.2 provides a general overview of dementia by explaining statistics of its prevalence, reviewing the risk factors increasing the chances of developing dementia, and discussing the most prevalent types of dementia, including AD, in more details, as well as describing what happens in the human brain as dementia develops. Section 2.3 gives an overview of all medical tests used in diagnosing dementia, as well as discusses their challenges and limitations. This is followed in Section 2.4 by a brief study of the most popular performance measures used to evaluate the CAD methodologies in diagnosing dementia. Section 2.5 highlights the most common features that have been used in state-of-the-art CAD methodologies to enhance early detection of dementia. Section 2.6 provides a comprehensive review of the CAD methodologies that have been implemented for early detection of dementia, and identifies their gaps.

2.2 Overview of Dementia

2.2.1 Incidence and Prevalence of Dementia

Over the past few centuries, many researchers have worked on developing the general notion of dementia. The Latin term "Dementatus", which means out of one's mind, is considered as the root of term "Dementia" [2]. According to [19], the physician Aurelius Celsius was the first person who used the term "dementia" as a medical term to describe delirium resulting from having got a fever during the first century A.D. By the fourth century A.D., the physician Oribasus conducted his first attempts to describe an etiology far from aging through creating a relationship between the signs of cerebral atrophy condition and the aging process [8].

In recent years, the prevalence of dementia has risen incredibly throughout the world. In England alone, more than 500,000 people had been diagnosed with dementia in 2011[15]. This number has significantly increased such that, in 2015, it has been reported that 46.8 million people worldwide were living with dementia and the projections indicate a significant increase which is estimated to be about 75.6 million by 2030, and will almost triple by 2050 to 135.5 million [56]-[57]. Total health-care costs for people with dementia amounted to more than 1 per cent of the global gross domestic product (GDP), or US $ 604 billion, in 2010 [7]. Based on the facts mentioned above, it can be concluded that the prevalence of dementia forms is expanding rapidly in most world regions. Generally, the number of cases with different types of dementia increases due to the exposure to various environmental and genetic factors that are discussed in the next section.

2.2.2 Causes and Types of Dementia

Dementia can develop as a result of a number of risk factors; some of these factors are advanced age, genetic disorder, traumatic brain injuries, and environmental factors [16]. Advancing age is the greatest known risk factor of dementia. The majority of dementia cases have been diagnosed in people over 65 years old [3]. As well, family history may play a role in increasing the likelihood of developing dementia through having dominant genetic mutations; abnormal changes in the sequence of chemical pairs that make up genes [58]. Furthermore, the process of developing dementia could be increased by environmental factors, such as drinking alcohol above the recommended level and excessive smoking [59]-[48]. In addition, exposure to head injuries frequently or suffering from atherosclerosis and hypertension could increase the risk of developing some types of dementia.

Dementia disorders have many forms whose signs are sometimes similar; nevertheless, some types of dementia are more popular than others. Alzheimer's disease (AD) is the most prevalent type of dementia, as can be noted from Figure 2.1. AD accounts for nearly 63 percent of all diagnosed cases of dementia and it has been estimated that almost five million people from various age groups in the United States of America have been diagnosed with AD in 2015 [5]. AD was discovered by Dr. Alois Alzheimer in the middle of 1906. It is defined as a neurological disorder caused by loss of brain cells that results in memory degeneration and cognitive impairment [60]. As illustrated in Figure 2.1, the second most common type of dementia is called vascular dementia (VaD). Recently, the percentage of patients with vascular dementia has risen gradually in which 10 percent of all cases diagnosed with dementia have been clinically diagnosed as having VaD [6]. Additionally, having abnormal characteristics from more than one type of dementia concurrently represents mixed dementia; which is the third prevalent type of dementia. Overall, many other forms of dementia have been clinically diagnosed, including Lewy-body dementia, Frontotemporal dementia, young onset dementia, Mild Cognitive Impairment (MCI), and some other rare forms [61]. As a result of the increased prevalence of AD and VaD around the world in comparison with other dementia types, these two types will be discussed in more depth through the rest of paper.

2.2.3 Understanding Dementia

Once dementia disease is triggered, whether by having any of the abnormal genes or through the exposure to any of the environmental factors, some abnormal changes take place in the human brain that affect its functionality. For example, plenty of abnormal structures

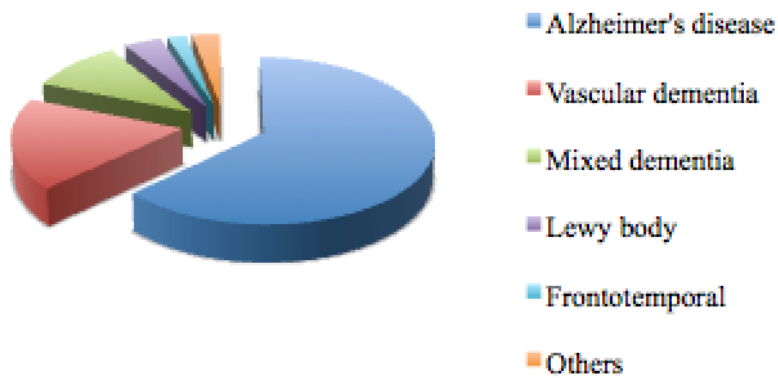

Figure 2.1: The prevelance of types of dementia.

called plaques and tangles form in the brain causing AD, as shown in Figure 2.2 [62], or the oxygen supply decreases to the brain causing VaD, which may be the result of a series of injuries to blood vessels in the brain [60], as seen in Figure 2.3 (D) [63]. Due to an abundance of these plaques and tangles in the brain, nerve cell connections are affected. As a result, these nerve cells may die [64]. Also, the human living brain has neurotransmitters-significant chemicals that are responsible for conveying signals among nerve cells. The amount of these chemicals decreases in the brain when dementia develops. Eventually, the symptoms of the disease start to emerge because of the occurrence of these abnormal changes.

Some symptoms of dementia like memory loss, personality change, disorientation, mood swings, and bad concentration are shared among its various forms while some others have

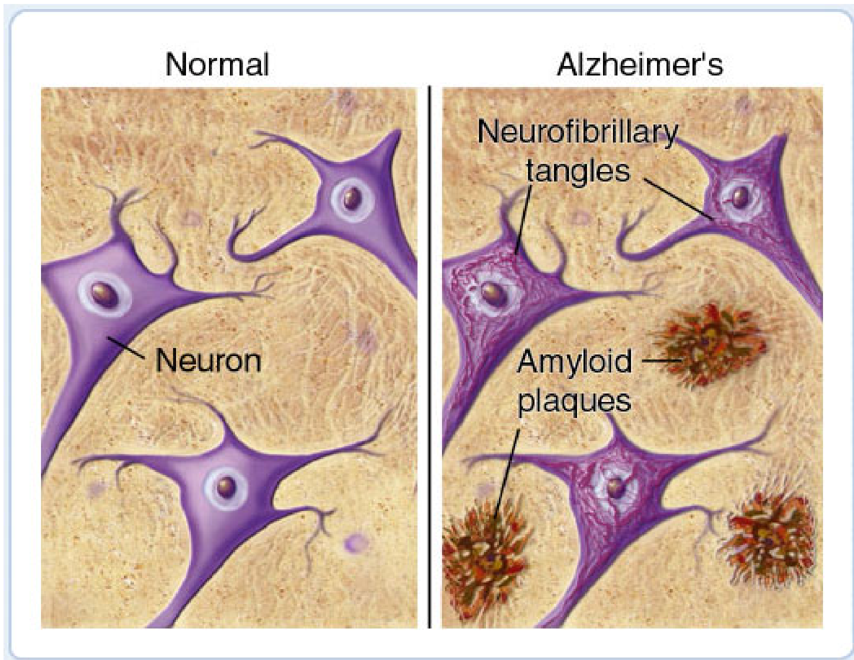

Figure 2.2: Difference between structures of normal and Alzheimer's diseased brains [62].

been clearly distinct to some types [65]. The diagram shown in Figure 2.4 clarifies the process that the human brain goes through as it develops dementia. When studying and investigating scans of normal brains and demented ones, many differential characteristics can be observed among them in terms of the cerebral cortex, hippocampus, and ventricles. First, as shown in Figure 2.3 (B) and (D), the cerebral cortex; the outer layer of the brain, referred to by arrow 1, looks smaller in a brain with dementia compared to in a healthy brain. In terms of hippocampus's size, referred to by arrow 2 in Figure 2.3 (B) and (D), it can been seen that the hippocampus is shrivelling up in a demented brain while it looks bigger in a normal brain. While the ventricle region, referred to by arrow 3 in Figure 2.3

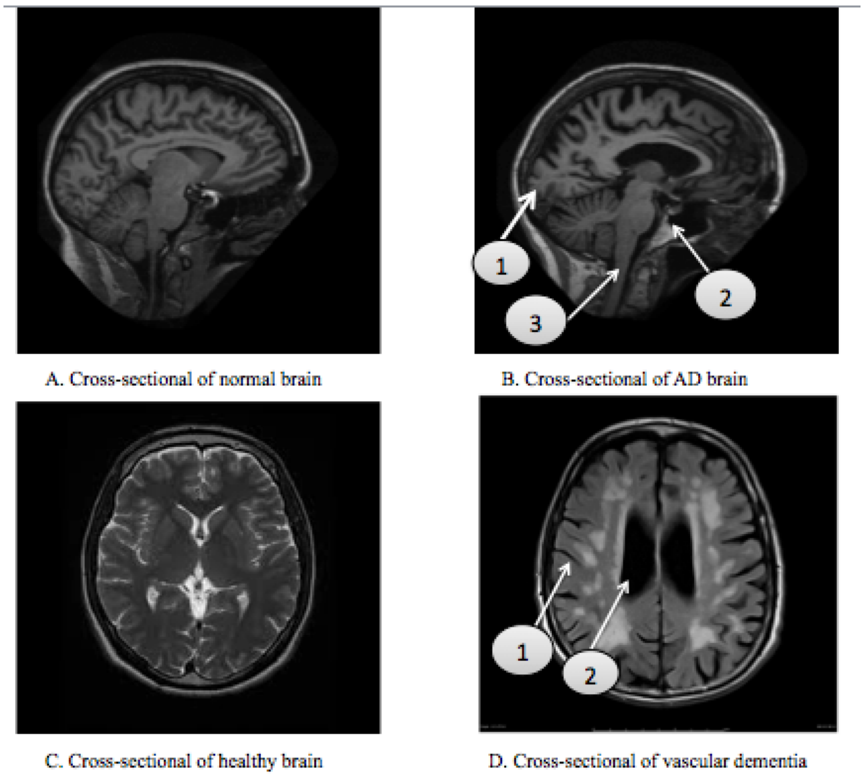

Figure 2.3: The differences between Normal brain, brain diagnosed with Alzheimer's disease and brain diagnosed with vascular dementia.

(B), looks enlarged in brain diagnosed with dementia compared to that in healthy brain.

2.3 The Diagnosis of Dementia

The process of diagnosing dementia is achieved by the examination of its symptoms and signs to identify the nature of the disorder. Diagnosis of dementia, especially in its early stages, is often complicated due to the absence of conclusive tests diagnosing the different types of dementia. As illustrated in Figure 2.5, the process of diagnosing dementia may

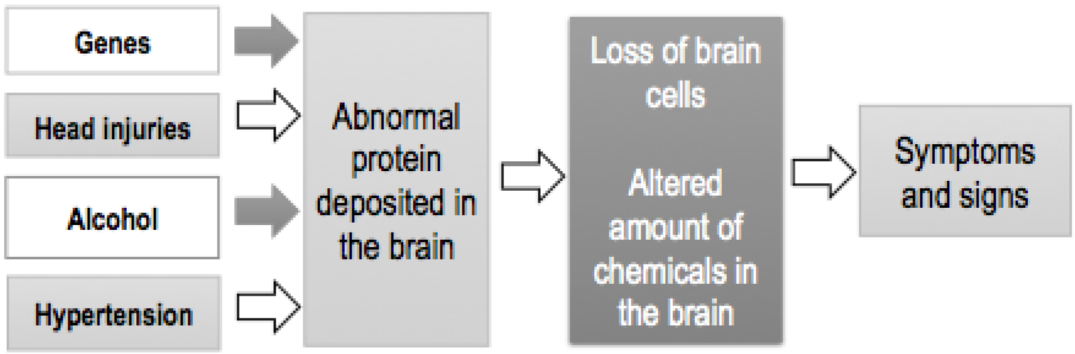

Figure 2.4: What happens in dementia?.

involve five different categories of exams that include: physical examinations, neurological examinations, neuropsychological tests, laboratory tests, and brain imaging techniques [5].

During the physical examination, the physicians are expected to review all medications used by the patients; and check their blood pressure, temperature, and heart pulses. As well, physicians are expected to inquire about dietary habits and nutrition, in addition to the use of alcohol, to identify the probability of other health problems that may cause strokes or brain tumors [66]. In contrast, through evaluating memory, reflexes, speech, eye movement, and muscle tones, neurological examinations are used to detect nervous system problems that are causing difficulties with thinking and behavior [67]. Patients may be subjected to laboratory tests used to collect blood or urine samples in order to detect physical problems affecting the brain, such as vitamin B-12 deficiency or an underactive thyroid gland. Also, patients could be subjected to a Mini Mental State Examination

(MMSE); a 30-points questionnaire in which a score of 23 or lower is indicative of cognitive impairment, in order to evaluate mental abilities like attention, memory, language, and planning [28]. In addition to all previously mentioned clinical tests for diagnosing cases with dementia, computer imaging is widely used in capturing scans of the human brain in order to rule out other brain disorders. Overall, it has been noticed that neuroimaging is more effective than clinicians and radiologists in the detection of dementia in its early phases.

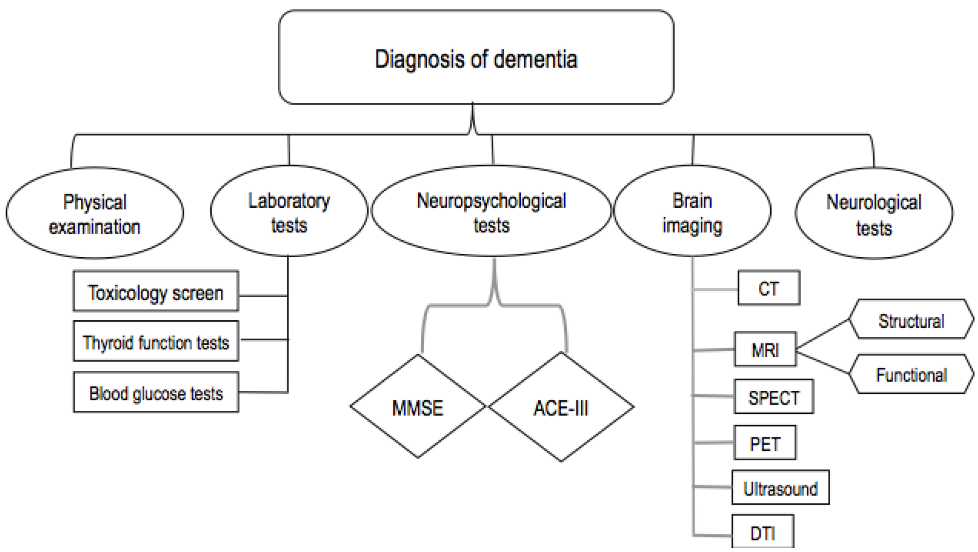

Figure 2.5: The set of common tests used for diagnosing dementia.

2.3.1 Brain Imaging Techniques for Diagnosis of Dementia

Neuroimaging aims to create high-quality images of structure and function of the nervous system inside the human brain to detect brain disorders by using any, or the combination, of medical imaging techniques, such as computerized tomography (CT), magnetic resonance imaging (MRI), positron-emission tomography (PET), and single proton emission computerized tomography (SPECT). These four imaging techniques are the most commonly used techniques in diagnosing dementia. The procedure of CT scans is used to produce cross- sectional images through taking a series of x-ray images of the brain from several angles and using computer processing techniques to fuse them. In addition, the MRI scan is a radiology technique that helps to show structure and metabolic function of the brain through using a magnet and radio frequencies [68]. Functional SPECT, that provides information about regional cerebral blood flow, has been found to be a valuable aid for early diagnosis of AD while PET provides information about Amyloid beta plaques deposition [29]. Overall, radioactive tracers are used during SPECT scans in order to track cerebral perfusion distribution of substance within brain tissue as well as these tracers disclose Amyloid beta plaques in the brain during PET imaging. Brain imaging outperforms all tests mentioned above in early detection of dementia in which it distinguishes among dementia and other disorders like tumors that may lead to symptoms similar to that of dementia. During diagnosing dementia, the use of neuroimaging techniques in achieving this purpose makes it possible to visualize the structure and function of the brain in a noninvasive way and see various changes in the brain [69].

Table 2.1 provides a comparison between these four imaging techniques in terms of possible negative impact, costs of the procedure, the quality of images, and the time spent in conducting the brain test. As shown in Table 2.1, all stated imaging techniques, except MRI, have an effect on the patient's body due to an involvement of ionization radiation [70]. In terms of the financial cost, PET is the most expensive technique among all the aforementioned radiology techniques [71]. Regarding the quality of the image, MRI provides high quality images with high-resolution, and more details about soft tissue and anatomic structures such as gray and white matter in the brain. Among all previously discussed imaging techniques, CT is the fastest technique in providing information during diagnostics procedure.

Table 2.1: Comparison of Different Imaging Modalities

Measures	CT	MRI	PET	SPECT
Cost	Low- Med	Low - High	Very High	High
Effect on body	Yes	No	Yes	Yes
Quality of image	Low	Very high	High	High
Examination time	Only 15 min	30-45 min	30 m-1 hour	30 m-2 hours

2.3.2 Challenges of the CAD of Dementia

Although diagnosing dementia by using brain imagery techniques yields a number of significant advantages, it has proven to be a challenging task [72]. The diagnosis of early stages of dementia further complicates the task as it encounters many challenging problems. The following list discusses the most prominent set of these challenges:

- Interactive visualization and navigation: During the scanning of a huge amount of data, the process of navigating and visualizing this large data collection is challenging. For each human brain, different cuts of the brain are taken and stored in a dataset. Each of theses cuts is scanned by Polarized Light Imaging (PLI). After interactively visualizing nerve fibers by using a 3-D navigator, the visualization of these nerve fibers merged with PLI scans needs to be stored in a memory; therefore, the memory of the Central Processing Units must be a very huge space to be able to store all datasets without being overloaded [21].

- Compression for storage or transmission: Most medical images need to be compressed using lossless compression in order to store the huge amount of images or transmit them. Data for each pixel of the image is valuable. Image compression is one of challenging problems in image processing since the compressed image could emerge with severe loss of image quality.

- Quality of image: The quality of medical images is calculated by measuring the degradation of the image. In fact, the quality of medical images relies on various factors, including noise, image contrast, and artifacts. Many researchers have worked on improving the quality of medical images. In particular, some researchers sought to increase the MR image quality by reducing image noise through an application of lossy compression as a noise reduction filter [30].

- Preprocessing and Segmentation: Image preprocessing and segmentation is key to standardize image quality, in terms of resolution, artifacts, and noise levels. Such preprocessing can be used for measuring and visualizing the brain's anatomical structures, analyzing changes in the brain, and delineating pathological regions. The accuracy of the CAD methodology may significantly depend on the accuracy of the preprocessing and segmentation technique used.

2.4 Performance Measures of CAD of Dementia

The performance of CAD of dementia in literature is assessed through the calculation of various measures, such as sensitivity, specificity, predictive value, reliability, and the accuracy of diagnosis in early detection of dementia. The rest of this section discusses these performance measures.

2.4.1 Validity

The validity of the available set of screening tools in identifying early dementia is measured by investigating the ability of these tools to differentiate between healthy people and patients with dementia as well as investigating the ability to differentiate between subgroups of patients based on type of dementia [13]. This validation is evaluated by computing sensitivity, specificity, and predictive values of diagnosing dementia.

- Sensitivity: The probability of diagnostic tests to identify the subjects who have dementia [9]. The CAD sensitivity can be calculated by the formula given in Equation 2.1, in which the value of a refers to the total number of patients having dementia as a positive result of the test while the value of c indicates to total number of people with dementia who tested negative. Thus, a + c refers to the total number of subjects with dementia.

$$Sensitivity = \frac{a}{a+c} \tag{2.1}$$

- Specificity: The probability of negative tests identifying subjects without dementia. The CAD specificity can be calculated by the formula given in Equation 2.2 [9]. Specificity measures the total number of patients the system classifies as not having dementia, referred to by d, compared to the total number of subjects without dementia, refers to by d + b in which b indicates to people without dementia who tested negative.

$$Specificity = \frac{d}{d+b} \tag{2.2}$$

- Predictive values: Predictive values (PV) are the sum of rates of people who are correctly identified as having dementia, i.e., true positives (TrP), and rates of people who are correctly identified as not having dementia, i.e., true negatives (TrN) [73]. PV is given by

$$PV = TrP + TrN \tag{2.3}$$

2.4.2 Inter-rater Reliability

Many studies sought to evaluate Inter-rater reliability of traditional early diagnosis methods of dementia. In fact, Inter-rater reliability is evaluated by measuring how many physicians diagnose the same samples of patients by using the same diagnostic test and got the same diagnosis results. CAD systems aim to support radiologists in the early diagnosis of dementia by providing preliminary interpretations of the medical images [31]. These CAD systems are responsible for analyzing the medical imagery and yielding a decision on their interpretation. Therefore, CAD algorithms need to preform with high reliability such that these algorithms should continue to produce similar results under consistent conditions over time.

2.4.3 Accuracy

The key performance measure for CAD is the accuracy of the image interpretation in differentiating normal controls from patients with cognitive impairment. The accuracy can be evaluated by a number of measures, some of them discussed below:

- Classification Accuracy: It is the percentage of correct predictions in identifying patients with early stages of dementia to the total number of samples. The classification accuracy can be calculated by Equation 2.4, in which the value of a + b + c +d refers to the total number of subjects.

$$Accuracy = \frac{TrP + TrN}{a + b + c + d} \tag{2.4}$$

- Confusion matrix: It is a two dimensional matrix where each row of the matrix represents an actual class while each column represents a predicted class. By evaluating the performance of a CAD algorithm against a test dataset with known ground truth diagnosis, each matrix cell refers to the number of subjects of the predicted class with respect to their actual class.

- Precision: It refers to the closeness of two or more measurements to each other. Precision is independent of accuracy. In order to get the high precision of dementia diagnostics, CAD systems provide consistent interpretations of medical images.

2.5 Brain Image Features For Diagnosing Dementia

Individual measurable properties of the human brain can be observed to discriminate healthy brains from those brains that are developing dementia. This section describes the most popular features in the human brain observed and measured by the state-of-the-art CAD methodologies for the early detection of dementia.

2.5.1 Voxel-Based Morphometry (VBM)

VBM is a neuroimaging analysis technique that has been developed to identify the focal differences in the structure of human brain. The VBM involves normalizing brain images and then segmenting it into gray matter and white matter that are smoothed using an isotropic Gaussian kernel. Finally, the differences between gray and white matter are analyzed [32]. VBM has been used to characterize a variety of diseases, including Down's syndrome, Parkinson's disease, and Alzheimer's disease.

2.5.2 Regions Of Interest (ROIs)

ROI in medical images is an area on a digital image that circumscribes a desired anatomical location in the human brain. The characteristics of the ROIs are analyzed and average parametric values are computed. Extracting morphologic properties of an ROI, including volume and dimensions of length, assists in distinguishing between normal brain and brain diagnosed with dementia [17]. Some researchers sought to calculate the volume of the ROI to identify the demented patients through use of Multi-Atlas based Label propagation (MALP) and the expectation maximization algorithm [33].

2.5.3 Hippocampal Regions Intensity Features

Hippocampus is one of the key parts of the brain that is responsible for memory. However, the hippocampal region may suffer from neuro-degeneration early as dementia develops. There are a number of reliable intensity features of the hippocampal region that determine the hippocampal body health. These features include the mean, kurtosis, skewness, standard deviation and wavelet-based texture features of the hippocampal region [34].

2.5.4 A Combination of Volume, Thickness, and Shape

- Hippocampal volume: Studies have identified the relationship between memory performance and hippocampal volumes [35]. The volume of hippocampus in the brain can be measured by the Multi-Atlas Segmentation (MLS) method and then using an expectation maximization algorithm [36].

- Cortical thickness: Cortical thickness represents the thickness of the outer layers of the cerebral cortex in the brain. It is one of the main commonly used features in the state-of-the-art CAD of dementia methodologies. The cortical thickness of the brain commonly determined on the basis of the brain's gray matter segmented in neuroimaging data [107].

- Shape features: The shape features could be computed by the use of one of the spectral shape descriptors, like ShapeDNA [37]. There are various types of shape features that can be computed, such as BrainPrint and shape differences between different joined regions within the brain [14].

- Volume and intensity of entorhinal cortex: Lot of models that have been developed recently for the detection of dementia spotlights changes in hippocampus and entorhinal cortex. The entorhinal cortex is a region of the brain, located near from the hippocampus. Patients diagnosed with dementia, especially those with AD usually suffer from atrophy of entorhinal cortex [35].

2.6 A Classification of State-of-the-Art CAD Approaches of Early Detection of Dementia

Our problem is to detect multiple types of dementia in their early stages using MRI scans. Significant effort has been made to develop an effective approach to aid in the identification of early cases of dementia. Recently, various state-of-the-art CAD methodologies for early detection of dementia have been proposed. These methodologies are classified with respect to the need for preprocessing, features used to describe dementia biomarkers, and learning techniques. Initially, this relevant literature can be divided into learning techniques, which require segmentation and feature extraction as preprocessing steps and techniques, which can extract features from input without such preprocessing. Fig. 2.7 shows the classification of the CAD methodologies for identification of dementia.

Before training some learning techniques, a meaningful representation of observation is obtained by using feature extraction techniques as preprocessing step and then classified into categories like NC or patients with dementia based on the extracted features. Here, the following techniques, which perform data classification after preprocessing it (see Table 2.6.1.3).

Other techniques operate on features extracted from images by itself and classify them as NC or patient with dementia. The classification of techniques also can be based on the features, which are used to classify a given image as NC or not. Table 2.6.2.2 provides a comparison between the key CAD methodologies in literature, not required preprocessing for detecting dementia.

2.6.1 Techniques Requiring Feature Extraction as Preprocessing step:

2.6.1.1 Support Vector Machine (SVM)

Numerous research efforts have proposed using Kernel Support Vector Machine (SVM) [80] as a classifier to discriminate NC from patients with dementia. In [1, 26, 35], researchers have applied Voxel-based morphometry (VBM) [50] using Statistical Parametric Mapping (SPM8) to analyze data, reduce a set of features and then learn the meaningful representation of data. These studies computed the grey matter densities from medical images to classify various stages of AD. On the contrary, Sarica et al. [18] proposed using FreeSurfer?s means to learn features from MRI scans. The extracted features included cortical and subcortical volumes, and hippocampus volume to distinguish between AD, MCI, and NC. In [28], the researchers have also focused on reducing the high dimensionality of input data before training SVM for classification and transforming data into a reduced set of features. Dimensionality reduction of the neuroimaging data was carried out using Principal Component Analysis (PCA), Partial Least Square, and Independent Component Analysis. Ramrez et al. study was conducted through combining discriminant image parameters as best input vector as well as using ROIs for classifying and selecting the discriminant slices - coronal and sagittal - of the brain. As a result, the accuracy that results from this algorithm could yield up to 90% accuracy [34].

2.6.1.2 Linear Discriminant Analysis (LDA)

A couple of papers has proposed Regularized Linear Discriminant Analysis (LDA) for purpose of multi-class classification of AD, MCI, and NC based on the extracted features from neuroimages. The aforementioned features include the changes of brain volume, the thickness of cerebral cortex, the score of hippocampal shape, and the score of hippocampal. On both works, the features were learned using FreeSurfer. The results of Sorenson algorithm show that the accuracy of the diagnosis of dementia in its early stages is 63% [37].

2.6.1.3 Logistic Regression

Some researchers have proposed using a regression for classifying subjects as NC, AD, or MCI. Franke and Gaser [4] proposed the use of relevance vector regression, whereas Smith et al. [71] proposed using logistic regression for data classification. a CAD algorithm for dementia based on the analysis of the volumes of ROIs from fluoro-2-Deoxy-D-glucose (FDG)-PET and MRI images. This algorithm produces the high-level accuracy that may be up to 94% for early dementia detection [36].

Table 2.2: State-of-the-art CAD Methodologies based on Feature Extraction as Preprocessing Step

Algorithm	Publication dates	Features	Classifiers	Reported Accuracy	Imaging techniques
Rocchi et al. [36]	2015	ROIs	SVM	94 %	FDG-PET and MRI
Ramrez et al. [34]	2013	ROIs and image parameters	RBF-SVM	90.38 %	SPECT
Sarica et al. [18]	2014	Volume and thickness	RBF-SVM	53.7 %	MRI
Sorensen et al. [37]	2014	Volume, shape, intensity, thickness	LDA	63 %	MRI
Abdulkadir et al [1]	2014	VBM	SVM	60 %	MRI
Smith et al. [71]	2014	Area, volume, and raw intensities	Regression	32.2 %	MRI
Wachinger et al. [16]	2014	Volume, shape, thickness	Generalized linear model	80 %	MRI
Bron et al. [26]	2014	VBM	SVM	92 %	MRI

Rana et al. [60]	2015	Whole brain volume	SVM	86.76 %	MRI
Franke and Gaser [4]	2014	VBM	Regression	90 %	MRI

2.6.2 Techniques Learning Features without Preprocessing Step:

2.6.2.1 Artificial Neural Network (ANN)

The methods presented on these papers performed dementia diagnosis based on applying Artificial Neural Network (ANN). According to [75], an optimal interpolative neural network (OINN) algorithm has been developed in order to classify between patients diagnosed with AD and others with VD based on training ROIs extracted from SPECT images. The researchers in [76] suggested to pretrain Artificial neural networks (ANN) to enhance the performance of CAD methods in discriminating among AD, MCI, and NC. The initial training of ANN is performed by initially projecting samples in dataset into key features and then the relationship between key projected features and classification labels is maximized by the centered kernel alignment (CKA).

2.6.2.2 Deep Learning Techniques

These papers have developed effective approaches that have used deep learning approaches to aid in the identification of early cases of dementia. DL techniques allow for initially extracting the significant features and then implement a classification procedure [73]. In [17], the researchers sought to differentiate between patients with AD and healthy people using a new approach that combines a sparse auto-encoder [41] in order to learn the features extracted from MRI images and the filters of a convolutional layer under sparsity constraints in addition to 3D convolutional neural network that are built to solve problems of image classification. As a result, the accuracy of correctly predicting dementia using 3D convolutional networks outperforms the accuracy of prediction with 2D convolutional networks. As well, a new contribution introduced in [33] also focuses on improving the accuracies of three suggested classifications: AD vs. NC, AD vs. MCI, and MCI vs. NC based on using a deep-learning-based feature representation that has been developed with stacked auto-encoder (SAE). SAE has been used to explore a latent representation of input by extracting features like the volume of grey matter tissue and intensity from MRI and PET, respectively. Consequently, the accuracy of correct prediction of two types of dementia has been enhanced based on the proposed model in which accuracies of 95.9% and 85% of diagnosing AD and MCI, respectively, have been obtained.

To summarize, most of studies that have been discussed above used MRI technique for the early detection of dementia as seen in Figure 2.6. In contrast, the use of CT imaging in identifying the early dementia has not been notable. This is attributed to the fact that although CT scans are accessible and low cost, this imaging technique cannot capture brain function. Support vector machine (SVM) has been the most popularly used classifier in the state-of-the-art CAD of dementia methodologies, as shown in Table 2.6.1.3 due to its ability to identify which class or group an object belongs to based on the value of linear combination of feature values. VBM and ROI characteristics have been the most prominently used feature to diagnose dementia. From Table 2.6.2.2, it is clear that algorithms that use deep learning techniques yield higher diagnosis accuracy for early detection of dementia compared to their counterparts. Figure 2.7 shows the classification of the CAD methodologies for the identification of dementia based on the need of preprocessing.

Table 2.3: State-of-the-art CAD dementia methodologies, where preprocessing step is not required

Algorithm	Publication dates	Features	Classifiers	Reported Accuracy	Imaging techniques

Payan and Montana [17]	2015	Full images	Deep learning techniques	89.11 %	MRI
Suk and Shen [33]	2013	Volume of gray matter tissue and intensity	Deep learning techniques	95.9 %	MRI and PET
Sarraf and Tofghi [75]	2016	grey matter and white matter	LeNet-5	98.84%	MRI
DeFigueiredo et al. [76]	1995	intensities	neural net-works	80%	SPECT

2.6.3 Gaps in the Research for a Literature Review

There is a number of limitations in state-of-the-art and there are many gaps that need to be addressed. This include that all state-of-the-art research focused on the diagnosis of multi-stages of AD severity ignoring other dementia types that can be as devastating or even more. Furthermore, the high dimensionality of neuroimages can hinder classification performance. Moreover, it can be difficult to distinguish between the various dementia syndromes given the overlap in many common clinical features across the dementia types;

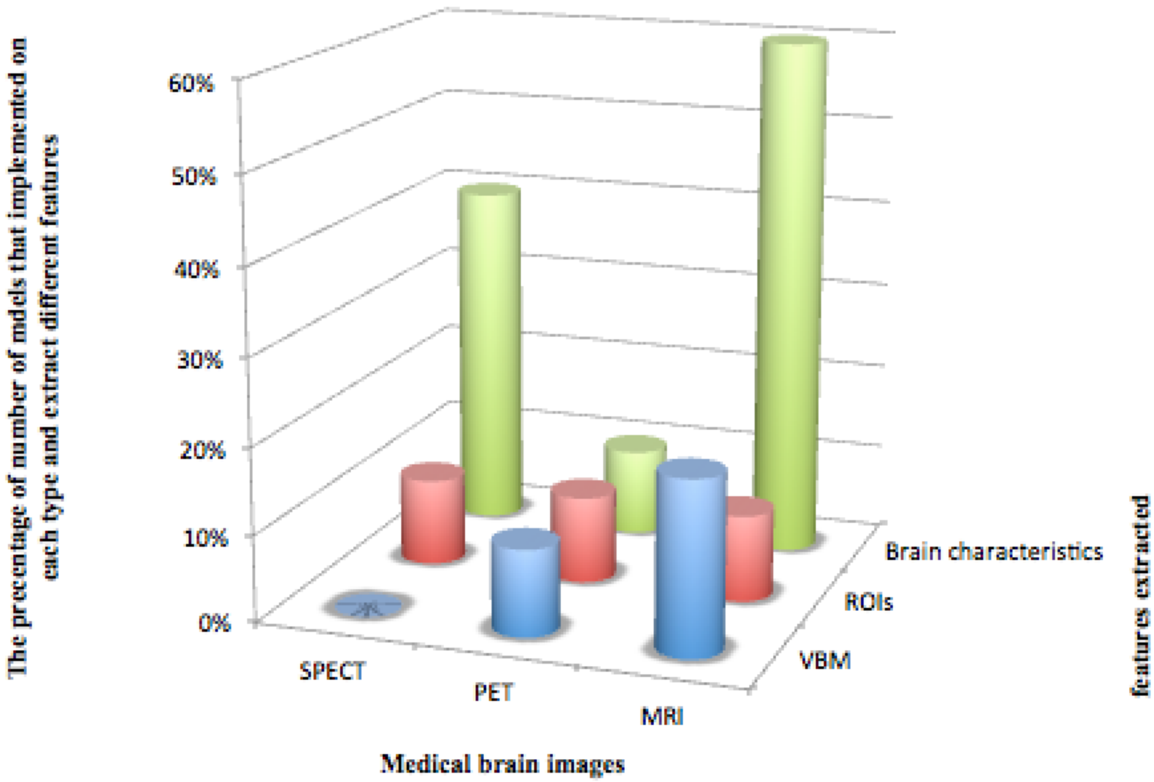

Figure 2.6: A comparison among different state- of- the- art imaging techniques used for diagnosing dementia.

eventually, this can lead to degrade classification performance.

2.7 Summary

Dementia has become one of the major causes of disability among older people worldwide. Dementia is very prevalent among elderly people, but is often overlooked even by skilled clinicians. This chapter provided a comprehensive survey of the state-of-the-art

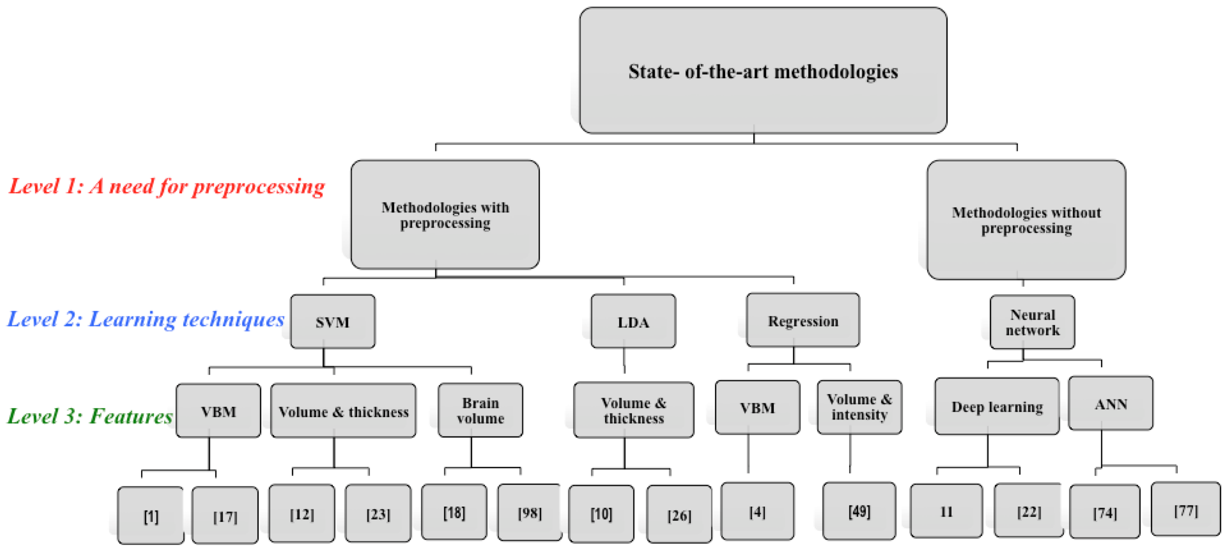

Figure 2.7: The classification of the state-of-the-art CAD methodologies for diagnosing dementia.

CAD methodologies for early detection of dementia. An overview of dementia has been presented investigating the causes of dementia, neuroimaging techniques used for its diagnosis, and the challenges of the early detection of dementia. The key characteristics of CAD methodologies have been investigated including brain features, learning techniques, and accuracy of detection. MRI scans are the most-used imaging technique in the CAD of dementia due to its high quality and its ability to capture brain functions.

A multitude of machine learning techniques have been investigated to enable early detection of dementia and provide a reliable classification approaches between different types of dementia. The most popular machine learning classifiers used include SVM and regression. However, the highest accuracy of classification has been obtained using deep learning techniques. Overall, deep learning architectures that been proposed as CAD approaches for diagnosing dementia, are yielding more promising classification accuracy than other classifiers. The state-of-the-art CAD methodologies focused on generating features associated with differentiating many severity stages of AD. In other words, they are lacking of acknowledgement of learning features needed for differentiating of other dementing conditions including VADs, Frontotemporal dementia, and other rare types of dementia.

In the following chapter, a new convolutional neural networks- based learning model has been proposed in order to learn set of features that is useful for detecting multiple types of dementia in their early stages.

Chapter 3

Multi-class Classification of Dementia

3.1 Introduction

Dementia is a rapidly progressing condition that affects the brain. Therefore, it needs to be detected as early as possible, preferably even before symptoms are evident. Early diagnosis of dementia considerably increases the chances of successful treatment. Hence, suitable care and support can be offered in a timely manner [51]. It is a well-known fact that structural brain imaging is playing a vital role in the identification of changes that occur in the brain associated with dementia. However, early diagnosis of dementia is challenging due to image noise and high dimensionality, as well as dementia biomarkers complexity. Accordingly, numerous researchers have devoted their efforts to develop new or improved Computer-Aided Diagnosis technologies based on medical neuroimages to detect dementia accurately.

Aforementioned state-of-the-art studies focused on differentiating various stages of Alzheimer' s disease (AD) severity, whereas other prevalent dementia types, such as vascular dementia (VaD) and frontal lobe degeneration that can be at the least as devastating have not been investigated by the CAD community. In an attempt to solve such limitation, this work aims to predict various types of dementia in early stages through training deep learning technique on MR neuroimaging data. We hypothesize that applying deep learning-based approaches is more efficient in early prediction of multiple types of dementia due to their capabilities to transform information layer after layer into higher-level representations, capturing discriminant biomarkers of different types of dementia.

This chapter presents a novel convolutional neural network-based CAD approach to aid in analyzing MR neuroimages to differentiate various types of dementia. The rest of this chapter is arranged as follows. Section 3.2 elaborates on the basic concepts of our approach, including an overview of three commonly used ML classifiers used for conducting experiments in this chapter. Section 3.3 presents a review of previous CAD approaches developed for detecting dementia. The experimental setup is detailed in Section 3.4. Additionally, the experimental results are reported in Section 3.5, with further discussion in Section 3.6.

3.2 Preliminaries

3.2.1 Support Vector Machine (SVM)

Support vector machine (SVM) is one of the best-known models in image classification. It is a supervised ML algorithm that analyzes data used for linear and non-linear classification. It separates a set of training data into their appropriate class labels. SVM maximizes a minimum distance between hyperplane and the nearest examples. SVM does not only supports binary classification, but also multi-class classification tasks through doing more enhancements, such as adding more parameters and constraints to optimize the separation among multiple classes [107]. Let $\{x_i, y_i\}$ represents a set of training data, whereby x_i is feature vectors, y_i denotes an array of class labels, and i={1,..,N} , where N denotes a number of training images. The linear classifier is represented by given Equation 3.1

$$y_i = argmax_{i=1,..,N}((W_i)^T \phi(x_i) + b_i) \tag{3.1}$$

where W_i represents weight matrix, b_i represents bias vector, and ϕ represents the function that maps training examples x_i into higher dimensional space.

3.2.2 Logistic Regression

Logistic regression is another commonly used ML methods for implementing classification problems. It provides more non-linearity on linear classifiers by using the sigmoid or logistic function. Logistic regression fits sigmoid function into training data { x_i, y_i} by minimizing classification error as given by equation 3.2.

$$p(y_i|x_i) = \sigma(\sum_{i=1}^{D}(w^T x_i)) \tag{3.2}$$

where σ is the sigmoid function, x_i represents training data, and y_i denotes class label associates with each data x_i. Multinominal logistic regression (MLR) model is an extension of the binary logistic regression model, used to predict more than two classes. MLR uses Maximum- likelihood estimation method to evaluate the probability of each instance. Given input data (x), the output is predicted by combined all input values linearly using weights and modeling them into binary values. Maximum- likelihood estimation method is used to estimate these weights from training data x_i.

3.2.3 LeNet-5

LeNet (LeNet5) is one of the shallow Convolutional neural networks. This network composed of two convolutional layers; two subsampling layers, one after each convolutional layer, two hidden layers and one output layer. In case of LeNet, logistic sigmoid is an activation function used for nonlinearity. Unlike the deeper CNNs, the activation function in case of LeNet is added on subsampling layers instead of after convolutional layers.

3.2.4 Deep Learning

Deep learning is an emerging and exciting area of machine learning research that has gained significant attention during the past several years [44]. This deep form of machine learning has improved a broad array of applications, including music generation [52, 53], medical image classification [54], pattern detection [79], and biometric identification [55]. Recently, it has been observed that deep learning can have a significant impact on the early prediction of dementia due to its a huge improvement in classification performance as well as its ability to learn high-level features. Because of the need to deal with massive amounts of and high dimensional data, as well as the difficulty in abstracting high-level features from data, a number of researchers have proposed the deep learning to retrieve more sophisticated features from training data. In 2015, certain researchers proposed CNNs in order to learn features extracted from MRI images for differentiating between patients with AD and healthy people [11,80]. In [108], AlexNet is finetuned and then pretrained with ADNI dataset to classify healthy controls and AD' s patients.

In comparison to conventional image classification techniques of machine learning, the deep learning allows for computational models to extract key features (learn multi-level representation) of inputs passing across many layers so that each layer learns features from the output of the previous layer as described in Figure 3.1. It learns a feature hierarchy automatically all the way from pixels into classification decision scores, whereas conventional ML techniques perform classification tasks based on using features extracted from raw pixels in preprocessing phase [78].

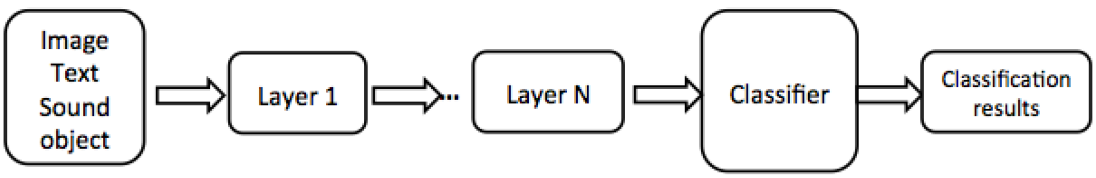

Figure 3.1: An overview of deep learning- based algorithm.

3.2.4.1 Convolutional Neural Network (CNN)

Convolutional neural network (CNN) is one of the most frequently used deep learning architectures, utilized in order to obtain good performance in various computer vision applications. CNN is widely used deep learning techniques in image classification due to its unique characteristics compared to other deep learning techniques. These Characteristics include the fact that its layers are fully connected. In addition, the max-pooling process down-samples image representation and reduces its dimensionality, thereby eliminating non-maximal values and reducing the computational cost. Typical CNNs are made of a set of different types of layers, that is one or more convolutional layers, a basic layer plus additional layers: the input layer, local response normalization layer, subsampling layer, and the output (fully connected) layer. Each of these layers takes a multi-dimensional array of numbers as input and produces an output that is another multi-dimensional array of numbers [81].

3.3 Related Work

Reasonable efforts have been made to aid in the identification of early cases of Alzheimer disease (AD) using medical neuroimaging modalities in AD. For instance, numerous work efforts have used Voxel-based morphometry (VBM) that computes the grey matter densities from medical images to classify the different stages of AD [1, 23, 34]. The second approach is based on analysis of some features of regions of interest (ROIs) from medical brain images including volume and cortical thickness [32, 35, 13]. In 2014, some researchers attempted to improve the accuracy of multinomial classification [39] that is provided by the combination of three binary classification models; NC vs. AD, NC vs. MCI, and AD vs. MCI [20]. Additionally, other studies proposed algorithms that evaluate changes in brain volume, measure the thickness of the cerebral cortex and compute the hippocampal shape score, and the hippocampal texture score from MRI brain images [36, 10].

Moreover, a regularized linear discriminant analysis classifier is used to differentiate between healthy and demented brains based upon the features mentioned above. In [28], researchers proposed a CAD for dementia methodology that distinguishes patients with Parkinson' s disease from NC by evaluating given samples using a calculation of sensitivity, MMSE, and specificity scores.

In recent years, several advances in research have been made to obtain accurate early diagnosis of dementia by the use of any of the deep learning approaches. Specifically, in 2015, some researchers sought to differentiate between patients with AD and healthy subjects using a new approach. Their method combined a sparse auto-encoder [43] to learn features extracted from MRI images and learn filters of a convolutional layer under sparsity constraints as well as convolutional neural networks that are built to solve problems of image classification [12, 44]. As well, another contribution has been introduced in [31] to improve accuracies of three suggested classifications: AD vs. NC, AD vs. MCI, and MCI vs. NC through implementing a deep learning-based feature representation that has been developed with stacked autoencoder (SAE) [20]. Moreover, SAE has been used to explore a latent representation of input by extracting features like the volume of grey matter tissue and intensity from MRI and PET, respectively.

Conversely, few research works proposed Artificial Neural Network (ANN)-based CAD approaches in literature to diagnose dementia. In 2016, Sarraf and Tofighi developed in [83] a CAD method that differentiates patients with AD from healthy controls by classifying features extracted from converted MRI images using the LeNet architecture proposed in [47]. Moreover, De Figueiredo et al. developed a new learning algorithm in order to differentiate between patients diagnosed with AD, patients diagnosed with VaD and healthy controls. They initially extracted relevant features and removed redundant ones from Single-photon emission computed tomography (SPECT) images. Therefore, Feature extraction has been applied by measuring of average intensity. Then, an optimal interpolative neural network (OINN) was used to analyze these features, and then predict a diagnostic output accuracy of dementia based on using k-means clustering algorithm [74]. As a result, the accuracy that results from this algorithm could yield up to 86% accuracy in differentiating between patients with VaD and NC. Subsequently, the work in [7] proposed an algorithm for pretraining ANN in order to enhance differentiating various stages of AD. First, the authors employed FreeSurfer software for analyzing MR images and producing feature vectors. Then, the projection matrix was created for feature vectors by implementing and maximizing the centered kernel alignment (CKA) for enhancing classification task.

This chapter proposes a novel deep neural network-based CAD method has been proposed to aid in analyzing MR neuroimages and then diagnosing dementia in its early stages. Unlike in [46], the CNN-based technique has been developed in this paper to accurately differentiate various types of dementia instead of focusing on differentiating multi-stages of

AD. Moreover, the suggested CNN based architecture consists of six convolutional layers and three fully connected layers while the architecture of CNN in [46] consists of only two convolutional layers.

3.4 The Proposed Method

In this work, a novel deep neural network-based CAD method is proposed to analyze MRI neuroimages to diagnose dementia in its early stages. This research work aims to differentiate multiple types of dementia based on MR imaging data. Our objective is accomplished through implementing a new two-stage CAD approach for detecting dementia, whereby the CNN is used to extract features from an input and then classify features into one of dementia types. The architecture of the proposed approach is illustrated in Figure 3.2. As shown in the figure, the proposed method initially preprocesses raw MRI neuroimages to remove image noise in addition to removing all bone and other portions surrounding the brain. Then, it pre-trains the CNN to extract latent feature representations from the segmented images and build robust CAD system. Finally, The learned features are then fed into the last fully connected layer of the CNN to identify the expected dementia diagnosis outcomes.

The various phases of the proposed approach will be further discussed in the following sections: In section 3.4.1, the image preprocessing techniques that are developed and used to improve the quality of grayscale images are presented, while feature extraction and classification are detailed in Section 3.4.2.

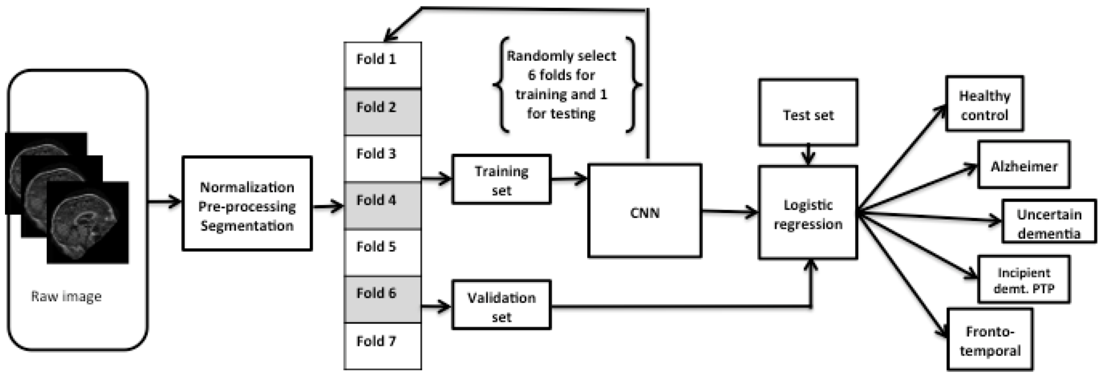

Figure 3.2: An architecture of proposed Convolutional Neural Network-based CAD algorithm for early prediction of various forms of dementia.

3.4.1 Image Preprocessing and Segmentation

This section discusses several image preprocessing techniques that are proposed to enhance MR scans before building the CNN. As MR images may be characterized with different sizes, All the raw images to be used as input are initially resized to be feed as input to CNN. In this research work, the images are resized to 176 $\times$ 176 pixels based on the sizes of the images in the OASIS dataset. The resized images are then preprocessed by digital spatial filtering [38] and image-adjustment techniques [43] to filter out various types of noise available in neuroimages. Overall, the preprocessing step plays a significant role in increasing the reliability of the features selected from the images and optimizing the images' quality.

In this research work, as seen in Figure 3.3, intensity normalization, Gaussian filter [52], and histogram equalization have been applied on the original MR neuro-images for preparing them of CAD system. Initially, these MRI images are normalized to the similar range of grayscale from 0 to 1because normalization reduces the range of intensity values simplifying feature abstraction process. Data is normalized by subtracting the mean and dividing by the standard deviation as illustrated by Equation 3.3

$$y = \frac{x - z}{\sigma} \tag{3.3}$$

where x is feature vector, z is the mean of that feature vector, and σ is its standard deviation.

Next, the Gaussian filter has been applied to remove or reduce noise from images with reduced size. The equation of a Gaussian function in two dimensions is

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}} \tag{3.4}$$

, where σ is the standard deviation of the distribution.

In addition, histogram equalization has been applied to improve the low contrast in grayscale image by adjusting image pixel intensities [18]. Suppose i denotes given MR neuroimage represented as $n_1 \times n_2$ array of pixel intensities, giving a range of possible values from 0 to 255, whereby zero tends to black color, whereas 255 refers to white color. Histogram of i is equalized by given Equation 3.5,

$$h = \frac{num\,of\,pixels\,with\,intensity\,n}{total\,num\,of\,pixels}, n = 0, 1, .., L - 1 \tag{3.5}$$

where L is the number of possible gray intensity values, often 256.

There are data instances that are incomplete and do not carry the data needed to address the problem. The presence of missing values in the training data often affects the performance of the model and, furthermore, it may lead to wrong and imprecise predictions. To avoid the performance issues and improve the classification accuracy, various imputation methods have been suggested in [82] in order to treat these missing values. In this research work, the missing data of numerical attributes is imputed by getting an average of the existing values in each attribute. Equation 3.6 is used for calculating the average of attributes with missing values. However, the most frequent value is selected for treating missing values for categorical variables in a dataset.

$$mean = \frac{sum\,of\,existing\,values}{total\,number\,of\,subjects\,existing\,values} \tag{3.6}$$

Segmentation also has an important role in analyzing medical images of the brain and allowing interpretation for the better diagnosis of dementia. Image segmentation is the process of dividing the image [33] to differentiate various regions, dissimilar characteristics in medical images. In this research, two different segmentation approaches are applied on MRI images of the brain to visualize the anatomical structures of the brain, and then analyze changes of that brain:

- Edge-based segmentation: Detection of a wide range of edges in an image is one of the important steps in retrieving the desired information from given images. Edges are lines identified on the basis of predicting the changes in various properties of the image like brightness, by using any edge detectors [24]. In this research work, Canny edge detection is performed to discriminate between gray and white matter tissues. This technique identifies edges that represent one of the significant features for analyzing brain images and prevent unconnected edges from continuing [38].

- Region-based segmentation: An image could be also segmented into regions. Each connected region is comprised of a set of adjacent pixels with shared characteristics and the same grayscale values. For applying this kind of segmentation in this research, the watershed transform algorithm is used to produce better segmentation results [37].

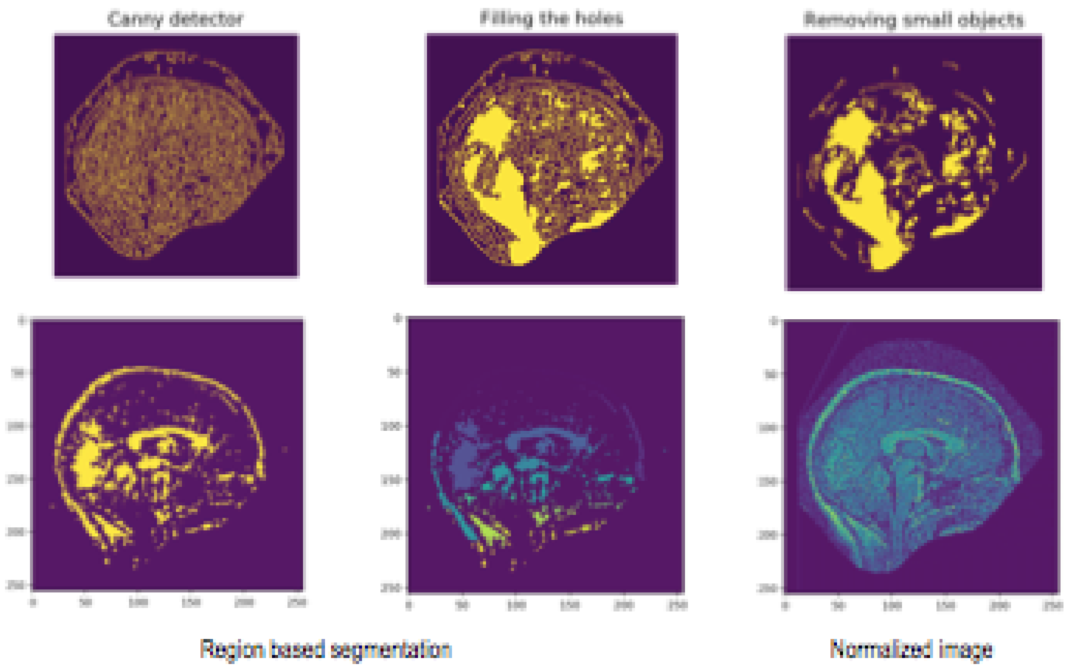

Figure 3.3: Preprocessing and segmentation results.

3.4.2 Feature Extraction and Classification

Accurate identification of the various types of dementia can be achieved by the use of the structural and functional biomarkers extracted from neuroimages. This chapter proposes to use the convolutional neural network (CNN) for extracting these high-level biomarkers and reducing the high dimensionality of large imaging dataset. Then, features from the top layers of the CNN are utilized for classification. As depicted in Figure 3.4, a deep CNN is trained to classify MR neuroimages in the dataset into the different five categories. The proposed CNN consists of six convolutional layers, which are followed by six max-pooling layers and three fully connected layers with a final 5-way softmax. These layers are stacked to build the full architecture of the CNN. In more details:

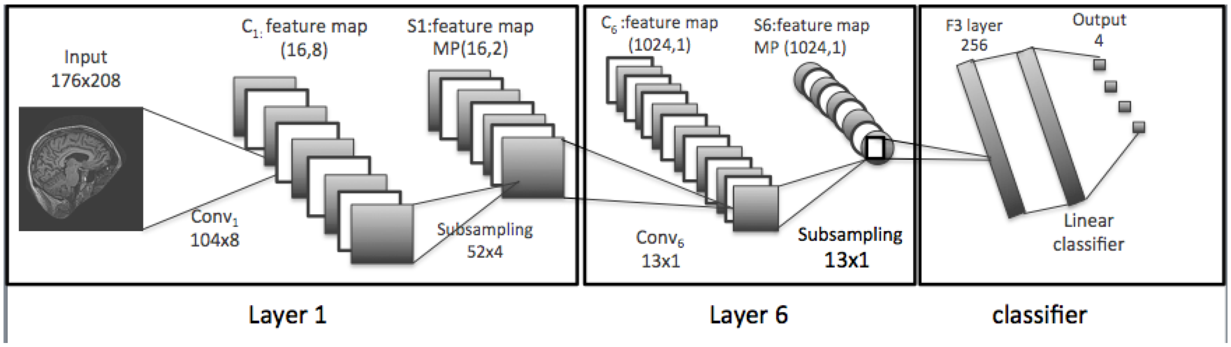

Figure 3.4: An architecture of the convolutional neural network for multi-class classification of dementia.

For the CNN input, the preprocessed 176×176 image pixels are reshaped into four-dimensional (4D) tensors. Each tensor consists of batch size, image width, image height, and the desired number of color channels, which is 1 in case of using grayscale images. The model input can be represented as follows:

$$x^i = [m_1, ...m_{30976}], i \in [1, 30976] \tag{3.7}$$

where i is the sample index, m_i is a column vector representing feature vector.

The extraction of the dementia features is achieved through a combination of the core parts of the CNN: the convolutional and pooling layers. The proposed CNN architecture includes six convolutional layers, which are followed by six max-pooling layers, with six local normalization layers, three fully connected layers, and two dropout layers. The proposed CNN can be more simply described if the input, the output, and the hyperparameters of the l^{th} layer of the CNN are considered to be identified by $x_l^j, y_l^j, and[W_l^j, b_l^j]$, respectively. The convolutional layer is denoted as C (k, s), which indicates that there are k kernels, called filters, each having a size of $s \times s$. Each filter composes a local patch of lower-level features into a higher-level representation by convolving the image matrix and then generating feature maps. The max-pooling procedure is identified with the use of an MP. During each convolutional layer, the stride is set to 1. Convolution is applied to the input image after reshaping it to a 4D tensor, and the MP is then applied on the convolution output to make the representation smaller and more manageable without the loss of a great deal of information representing feature vector. Dropout method is used to reduce overfitting in fully connected layers of the CNN. For first to seventh layers, the Rectified Linear Unit (RLeU) function is used as an output function to speed up learning procedure, introduce sparsity effect on the network as well as to avoid saturation issues. The output of the l^{th} convolutional and pooling layers can be written as:

$$o^j = MP(\sigma(W_l^j \times x_l^j + b_l^j)), j \in [1, k] \tag{3.8}$$

where σ is the activation function, and j is the index, considering multi-convolutional filter in the convolutional layer. Overall, the architecture of CNN can be described as: C (16, 8) - MP- LRN- C (32,4) - MP- LRN- C (64,2)- MP - LRN -C (128,1)-MP- LRN- C (256,1) - MP-LRN- C (1024,1) - MP - LRN.

The learned features that are outputted by the last convolutional and pooling layers are concatenated into a dense vector, composing the most high-level features of the input. Finally, the vector is transformed into outputs through the fully connected layer, in which the number of neurons in last fully connected layer is a number of types of dementia that can classify each instance of a dataset into one of the different categories of dementia. For optimizing the neural network, a back propagation algorithm can be used in order to train it to minimize cost and accelerate the convergence rate. In the proposed approach, a stochastic gradient descent algorithm, a popular back propagation algorithm, facilitates the achievement of the CNN training goal by adjusting its hyperparameters, including weights, learning rate, and biases. The details of the proposed CNN are listed in Table 3.1.

3.5 Experimental Results

The effectiveness of the proposed CAD method is evaluated by measuring some certain performance metrics, and then comparing the performance of the proposed approach to the performance of three state-of-the-art approaches, namely, Support Vector Machine

Table 3.1: Convolutional Neural network structure and hyperparameters adopted in our model.

No.	Layer type	Filter size	Stride	Output size	Nonlinearity
1	Input	–	–	176x176x1	–
2	Convolution	4x4x16	2	44x44x16	ReLU
3	Convolution	2x2x32	2	22x22x32	ReLU
4	Convolution	2x2x64	1	11x11x64	ReLU
5	Convolution	1x1x128	1	11x11x128	ReLU
6	Convolution	1x1x256	1	11x11x256	ReLU
7	Convolution	1x1x1024	1	11x11x1024	ReLU
8	Fully-connected	–	–	128	ReLU
9	Fully-connected	–	—	256	ReLU
10	Fully-connected	–	—	5	ReLU

(SVM), logistic regression, and convolutional neural networks with different depth.

3.5.1 Data Description

The performance of the proposed model has been evaluated on publicly available MR scans from the Open Access Series of Imaging Studies (OASIS) [97]. A classification of dementia was tested on 738 MR images collected from 74 different subjects. These images were taken from a dataset chosen from among many datasets available on the OASIS. OASIS is a multi-center project that consists of many brains' MRI datasets, collected from various centers, including Washington University Alzheimer' s Disease Research Centre, Howard Hughes Medical Institute, the Neuroinformatics Research Group, and the Biomedical Informatics Research Network.

Overall, each patient record consists of at least three cross-sectional brain scans. Figure 3.5 shows an example of three slices of MR brain images for each subject. Table 3.2 shows details of the dataset used for evaluating the proposed model' s performance. With regard to the general eligibility criteria in OASIS, subjects were in the age of between 18 and 96. General inclusion/exclusion criterias are as follows: All healthy control subjects had a clinical dementia rating (CDR) of 0 and MMSE scores between 24 and 30 (inclusive). On the other hand, subjects diagnosed with dementia had a CDR of at least 0.5 and MMSE scores between 20 and 26.

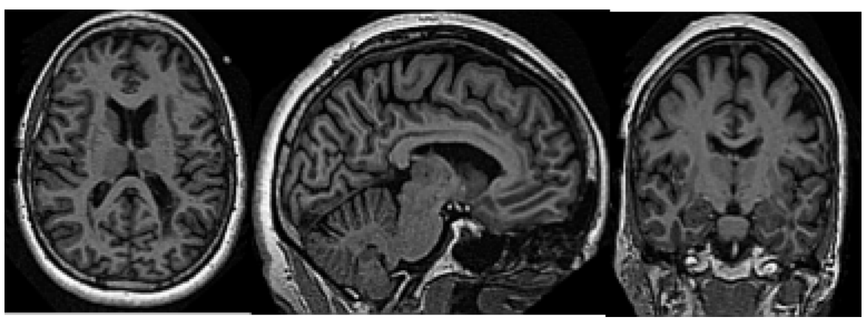

Figure 3.5: Three slices of patients' s brain with dementia.

3.5.2 Performance Evaluation Metrics

To evaluate the proposed model' s performance, several evaluation metrics described above in Chapter 2 have been calculated, including classification accuracy, sensitivity, and specificity. To do this calculation, true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs) have been initially measured as follows:

Table 3.2: Data description.

Attribute	Description	Values	Num. of subjects
Age	Age in years	18-25 26-45 46-65 66-96	45 8 16 40
Sex	Gender of patient	Male Female	45 68
CDR	Clinical dementia rate	0 0.5 1	89 16 8
MMSE	Mini-mental state examination	18-23=demented 24-30=NC	
nWBV	Normal of whole brain volume	0.644-0.797=demented 0.645-0.893=NC	
Sequence	MRI image sequence	3D T1-weighted images	

- **True positives (TPs):** The number of cases correctly predicted to have dementia.

- **True negatives (TNs):** The number of cases incorrectly diagnosed to have dementia.

- **False positives (FPs):** The number of cases correctly predicted to be normal controls.

- **False negatives (FNs):** The number of cases incorrectly diagnosed to be normal controls.

1. **Accuracy:** Equation 3.9 is used to calculate accuracy, which represents how often the proposed learning model is correct in its prediction of multiple types of dementia after measuring TPs, TNs, FPs, and FNs.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.9}$$

2. **Sensitivity:** Sensitivity is calculated by the fraction of relevant instances that have been retrieved over the total relevant instances in the image.

$$Sensitivity = \frac{TP}{TP + FP} \tag{3.10}$$

3. **Specificity:** Specificity is calculated by the fraction of those without dementia who will have a negative test result.

$$Specificity = \frac{TN}{TN + FN} \tag{3.11}$$

To evaluate the performance of the proposed model making predictions of various dementia types on new unseen data, seven-fold cross validation is performed on the MR images. Cross-validation reduces the chance of overfitting that occurs due to model complexity, thereby allows a good estimation of the proposed method's accuracy of proposed

CAD method in predicting multi-type dementia. After getting the preprocessed dataset, seven random datasets are produced to repeat training and testing of the model in which 60% of MRI images are assigned to the training set, whereas 40% of images are used for testing the algorithm. The comparison among the proposed method and some state-of-the-art models, including Support Vector Machine (SVM) [23], logistic regression and the LeNet-5 in [83] is made after training and testing them on the same MRI dataset in terms of following evaluation measures.

3.5.3 Classification Results

- **Investigating CNN depth for the proposed CAD approach:**

  In the designing phase of the CNN for recognizing dementia in MR neuroimages, some empirical experiments were conducted to select and tune various CNN parameters, including the number of hidden layers and the number of nodes in the CNN architecture as these parameters have profound effects on the system.These experiments were conducted at a specific learning rate for stochastic gradient descent that has been set to a value of 0.01. As well, the models were trained for 2000 epochs. Initially, I built the CNN consisting of two convolutional layers, followed by two maximum pooling and one fully connected layers. The results show that the proposed method is correctly predicted multi-types of dementia with an accuracy of 38.59%.

Then, the CNN has been developed with adding more hidden layers, so that it consisted of four convolutional layers, followed by four maximum pooling and two fully connected layers. From the results, it has been shown that an error is still decreasing, and the accuracy increases up to 50.87%. Finally, additional hidden layers have been added into the proposed CNN, in which two additional convolutional layers and two max-pooling layers, as well as one more fully connected layer, have been added to the proposed model to enhance classification performance.

The experimental results indicate that the classification accuracy increases from 50.87% to 74.93%. After that, it has been noticed that the error stop decreasing when the number of convolutional layers was increased than six layers. Table 3.3 presents the detailed results of an impact of the CNN with different depth on identifying of various types of dementia.

Table 3.3: A comparison between the proposed CNN-based approach and other approaches with different depth of the CNN.

Depth of CNN	Accuracy	Specificity	Sensitivity
Two convolutional layers	38.59%	33.33%	77.2 %
Four convolutional layers	50.87%	13.56%	**85.53%**
Six convolutional layers	**74.93%**	**46%**	63%

- **A comparison with State-of-the-Art Methods:**

  In order to evaluate the efficacy of the proposed CAD approach in predicting multi-type dementia in its onset stage, we compare its performance with the performance of three state-of-the-art techniques including SVM proposed in [23], logistic regression, and LeNet-5 in [83] in terms of measuring classification accuracy achieved by network, sensitivity, and specificity. However, the clinical diagnosis was used as the target during training and testing procedures. It is a categorical scale quantifying various types of dementia. As a result, Table 3.4 shows that the results of the proposed approach in classifying of five dementia types outperform the results of other compared ML models on testing data regarding accuracy and specificity while the sensitivity of the proposed model is worse than the sensitivity of others.

Table 3.4: Multi-label classification results for multi-type of dementia detection.

Machine learning classifier	Training accuracy	Testing accuracy	Specificity	Sensitivity
SVM [23]	67.01%	70.09%	33.33%	77.2 %
Logistic regression	67.04%	68.36%	13.56%	**85.53%**
LeNet-5 [83]	61%	70.04%	31%	74%
Our model	**70.97%**	**74.93%**	**46%**	63%

3.6 Summary

In this chapter, a new computer-aided classification system for differentiating the various types of dementia in its early stages has been built and tested. This CAD method combined convolutional neural networks with logistic regression. Also, a comparison of the performance of the proposed model and three different models was conducted.The results of this study indicate that the hypothesis that early diagnosis that is based on classification of feature extracted from MRI neuroimages is more accurate than early diagnosis that is based on classification of raw image pixels intensities has been accepted so that the proposed model yields higher classification accuracy compared to that using Logistic regression by 6 percent and that using LeNet-5 or SVM by 4 percent of classification accuracy.

These investigations about the performance of our model could be improved in the future studies. For example, the first convolutional layer could be pre-trained with autoencoder. In addition, 3D convolutions could be used instead of 2D convolutions to boost the performance of the model and enhance its prediction.

The next chapter proposes pretraining the first convolutional layer of CNN with low-dimensional features learned by sparse autoencoder (SAE) to boost the performance of the model and enhance prediction results of dementia. The learned features by the SAE and that obtained from the popular PCA approach, are then used to train a Linear Discriminant Analysis (LDA) and a Logistic Regression classifiers to compare their prediction accuracy.

Chapter 4

Feature Abstraction For Enhancing Classification Performance

4.1 Introduction

Due to the importance of early diagnosis for prevention or treatment of dementia, Computer-Aided Diagnosis (CAD) methods are deemed essential to enable early identification of dementia. Recently, neuroimages have become an important tool in the diagnosis of dementia and subsequent prediction of its progression. Neuro-images often suffer from the curse of dimensionality that can make them difficult to visualize and classify [84]. In fact, high dimensionality of images has a negative impact on the performance of Machine Learning (ML) classifiers during the training phase on these images, such that overfitting increases due to increased data complexity. To address this concern, feature-abstraction techniques have been developed in order to reduce and avoid the effects of the high dimensionality, enhance the accuracy of generalization, and then enable more robust systems. The feature abstraction is a technique that collects relevant features and ignores irrelevant or redundant ones from data, without the loss of much information can be developed to improve classification performance. State-of-art studies for diagnosing dementia have been implemented based on features extracted to differentiate between multi-stages of the severity of Alzheimer' s disease (AD). However, the feature abstraction to discriminate between multi-types of dementia has not yet been explored.

The main novelty of our approach is to extract key features from different Region of Interests (ROIs) and use them to differentiate between normal controls (NC), individual patients with AD and patients with Frontotemporal lobe dementia (FTD) using Sparse Autoencoder (SAE). The proposed method is verified by using Linear Discriminant Analysis (LDA), logistic regression and Convolutional Neural Network (CNN) in coupled with logistic regression (CNN+ regression) classifiers, with both the SAE-based low dimensional dataset and with PCA driven representation for comparison.The remainder of this chapter is organized as follows: Section 4.2 briefly reviews the definitions and notations. Section 4.3 is an overview of the most popular techniques for improving the dimensionality of input space to fit sample size. Section 4.4 introduces the CAD methodology proposed for producing the high-level representation of neuroimaging data so as to extract the key features that identify multi-types of dementia. Then, these features are compared to those extracted using the popular Principal Component Analysis (PCA) through training several types of dementia classification methods. The experimental results are discussed in section 4.5. Finally, Section 4.5 concludes this study.

4.2 Preliminaries

4.2.1 Principal Analysis Component (PCA)

PCA is probably the most common multivariate statistical techniques and one of the most commonly used techniques for reducing the high dimensionality of sets of data because it allows compressed images to be produced without loss of much important information. By applying PCA to the dataset, it finds principal variables for the given dataset by measuring the eigenvectors of a matrix and then listing them by their eigenvalues. The centered data can then be projected onto these principal axes to yield principal components (”scores”) obtained from singular value decomposition of given dataset. For the purposes of dimensionality reduction, one keeps only a subset of principal components and discards those with a smaller variance. Furthermore, it represents a pattern of similarity among probabilities and variables by displaying them as points on the map [94].

4.2.2 Linear Discriminant Analysis

LDA is one of the popular techniques for data classification and dimensionality reduction. Dimensionality reduction can be achieved by feature extraction [95]. The primary difference between LDA and PCA is that PCA does more feature extraction, whereas LDA does more data classification as well as high-level feature extraction.

For reducing the dimensionality using LDA, consider $(x_i, y_i)_i = 1^n$ represents a set of data points in the given dataset, where $x_i \in R^d, y_i \in 1, 2, ..., c$ represents class labels of i^{th} data point. Multi-dimensional data matrix is denoted as $X = [x_1, x_2, ..., x_n]$.

4.2.3 Sparse Autoencoder

Autoencoders (AEs) have been recently applied as a preprocessing step, with success for various computer vision tasks. An AE is an Artificial Neural Network (ANN) that can be carried out for reconstructing input data to fit into lower dimensional space [87]. The architecture of autoencoders, as seen in Fig. 4.1, is made up of a set of layers: the input layer, the output layer, and one or more hidden layers. Initially, the input layer represents an encoding function *f (x)*, mapping given data $x \in R^{numinput}$ to compressed data $h \in R^u$.Then, the hidden layer that represents an identity function measures the amount of information loss between the compressed representation of data and the decompressed representation. Finally, h is decoded to get $x* \in R^{numinput}$, which is a representation of the input reconstructed by applying a decoding function *g (x)* at the output layer. The input and output layers have the same number of neurons [11].

4.3 Related Works

Medical image segmentation and feature extraction for the early detection of dementia from medical imaging modalities are very important processes for deciding the right therapy at the right time. Significant effort has been made to develop an effective approach to aid in the identification of early cases of dementia. While dimensionality reduction and feature extraction are still crucial, various recent CAD methodologies for early detection of dementia have focused on image classification without reducing image dimensionality. For example, the researchers in [21] proposed a CAD for dementia methodology that distinguishes patients with Parkinson's disease from normal controls by evaluating given samples using a calculation of sensitivity, MMSE, and specificity scores. As well, Rocchi et al. [25] proposed an approach with some enhancements to differentiate between patients diagnosed with dementia and healthy controls by extracting ROIs from medical brain images.

High dimensional neuroimaging data has a lot of redundant and irrelevant features. For this reason, a number of researchers have focused on reducing the high dimensionality and obtaining better features by applying dimensionality reduction techniques in order to extract features from medical images and so detect and classify different stages of AD.

Davatzikos et al. [85] proposed using Linear PCA to reduce input's dimensionality and extract features by analyzing medical images of the brain.The extracted features are then represented as a set of orthogonal variables called principal components that are classified for differentiating between AD, NC, and FTD. On the other hands, Sorenson et al. [26] and Ramirez et al. [23] suggested Linear Discriminant Analysis (LDA) for extracting key features, denoting dementia biomarkers to enhance classification accuracy between healthy and demented brains. These features include the whole brain volume, the thickness of the cerebral cortex, the hippocampal shape score, or the hippocampal texture score, extracted from MRI brain images.

Several recent advances in research to extract the significant features of neuroimaging data have been made by the use of an autoencoder. Some researchers sought to differentiate between patients diagnosed with AD and healthy people by proposing a combination of a sparse auto-encoder (SAE) [27]. This SAE learns features extracted from MRI images and then the output is fed into convolutional neural networks to solve problems of image classification [11, 80]. As well, Liu et al., and Suk and Shen [22,86] proposed an AD classification methods based on a deep learning-based feature representation that has been developed with Stacked Sparse Autoencoders (SSAEs) [20]. SSAEs has been used to explore a latent representation of input by extracting features like the volume of gray matter tissue and the intensity from MRI and PET, respectively.

Thus, the existing literature indicates that the field of neuroimaging is well situated to provide a powerful tool for diagnosis of dementia. In comparison to the works in [11,80], instead of focusing on distinguishing multiple stages of AD, the proposed approach focuses on learning high-level information about brain's structure for differentiating among multi-types of dementia. The differentiation between the gray and white matter voxels has been suggested to generate high quality and detailed input data of SAE. Unlike dava and others who use the linear techniques to extract features [85], this chapter proposes to use a nonlinear technique for learning a nonlinear combination of features showing dissimilar properties of brains with AD, FTD, and NC.

4.4 Proposed Method

This research study aims to accurately classify multi-type dementia with lower computational costs and lower storage space. To achieve these objectives, SAE is proposed for learning high-level features associated with several types of dementia using the pixel intensities of MRI neuroimages. As seen in Figure 4.1, raw MRI images are initially carefully preprocessed and segmented, and then reconstructed with SAE for extracting the relevant features. Finally, these features extracted are tested with several ML classifiers to evaluate classification performance of the proposed method in early prediction of many types of dementia.

Image preprocessing and segmentation are detailed in Section 4.4.1 and 4.4.2, respectively. In section 4.4.3, the architecture of SAE is presented. Section 4.4.4 describes different ML algorithms applied for the classification task.

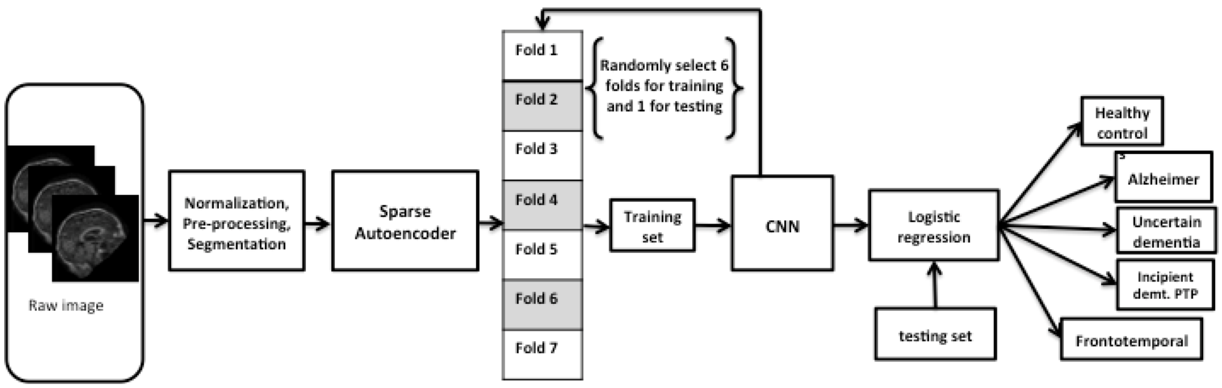

Figure 4.1: The architecture of a spare autoencoder-based CAD approach proposed to learn latent feature representation from MRI scans.

4.4.1 Image Processing and Segmentation

This section discusses several image-preprocessing techniques proposed for enhancing the MRI scans prior to the building of the CNN. With the raw MRI images as input, preprocessing involves converting the image format, resizing the images, normalizing the intensity, and stripping representations of the skull and all non-brain tissues. This is accomplished by importing MRI neuroimaging data into FreeSurfer software. After rescaling MRI images into a unified size, 176 pixels x 176 pixels, the downsize of images in the dataset, we initially convert MR neuroimaging data into a NIFTI format, followed by performing an initial intensity correction, that is followed by an automatic brain mask creation. The creation of brain mask removes the portions that reflect all non-brain tissues using the watershed transform algorithm. Figure 4.3 shows MR neuroimages after removing the skull and all non-brain tissues. Later, the images are segmented according to whether they depict cerebrospinal fluid (CSF), gray matter, or white matter as seen in Figure 4.4.

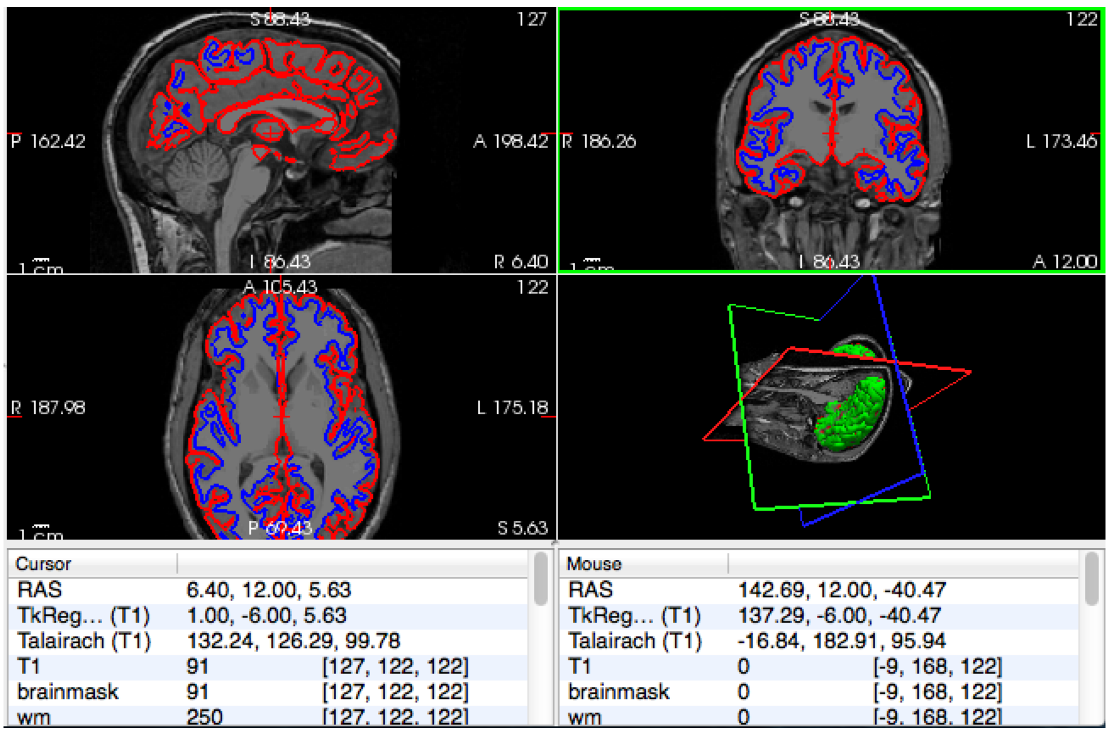

Figure 4.2: The raw images are imported into FreeSurfer software that converts images' format into NIFTI format.

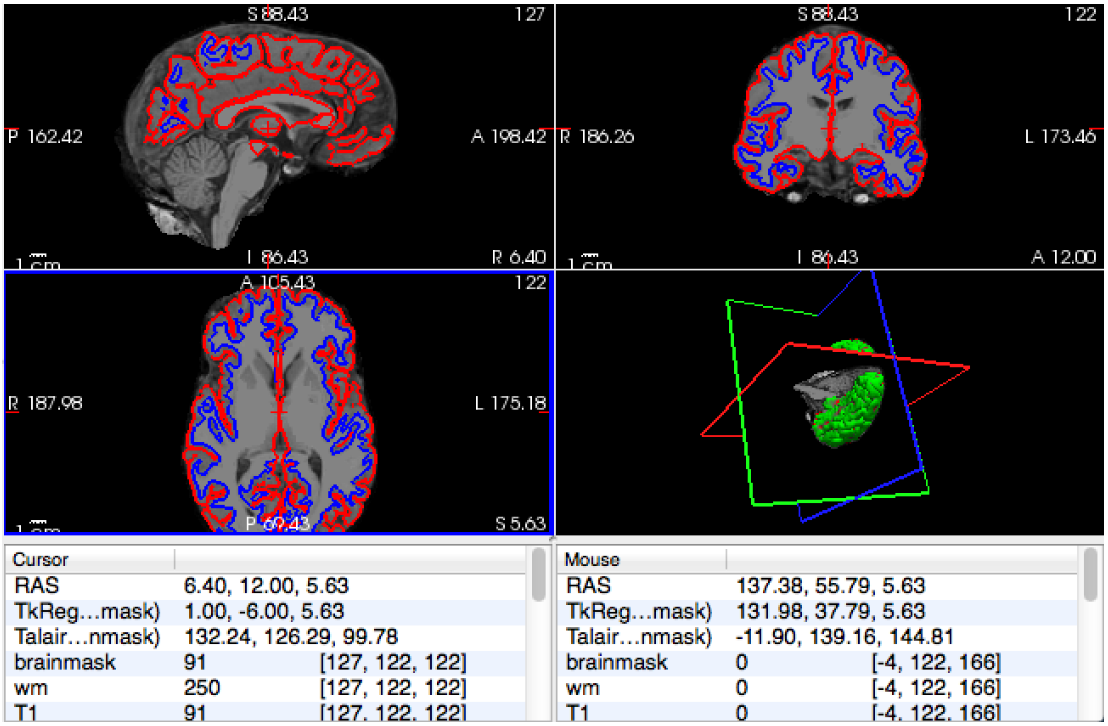

Figure 4.3: MRI images after removing the skull and all non-brain-tissues.

4.4.2 Feature Abstraction and Dimensionality Reduction

- Sparse Autoencoder: In this study, input xi represents pixel intensity values extracted from MRI brain images where $x_i = x_1, x_2, ..., x_{30976}$. Overall, the encoder function initially converts the input x to compressed form $h \in R^u$ that fits in limited spaces u, where u represents the number of hidden units as in Equation 4.1.

$$h = f(Wx + b) = W^*x + b^* \tag{4.1}$$

segmentation.png

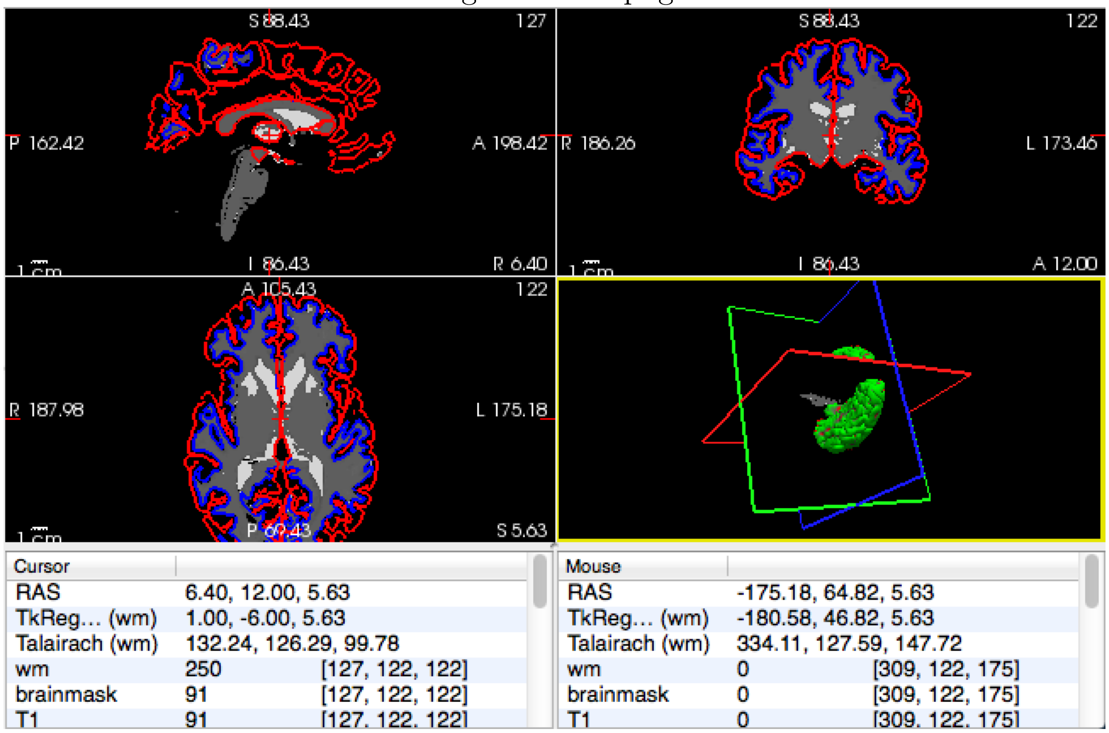

Figure 4.4: Segmentation of images into the white-gray matter tissues.

, where $W \in R^{u*n-input}$ and $b \in R^u$ represent the matrix of weight and the vector of biases of encoder function, respectively. Also, f is a sigmoid function.

Finally, the decoder function reconstructs the input x by taking the hidden representation of x as input to get x^*, which denotes relevant features for the dementia classification task. Equation 4.2 is used to apply the decoder function.

$$x^* = g(W^*x + b^*) = W''x + b'' \tag{4.2}$$

, where g is the identity function due to real-valued input as well as $W^* \in R^{u*n_input}$ and $b^* \in R^{n_input}$ denote the weight matrix and the bias vector of decoder function.

In addition to restricting AE by reducing the number of hidden units, the sparsity constraint pc, given by equation 4.3, is forced by AE on the hidden units u in order to make them produce a better low-dimensional representation of input x. To achieve this, the activation of each hidden unit, denoted as a, is averaged to be close to a given sparsity parameter.

$$pc = \frac{1}{n\_input} \sum_{i=1}^{n\_input} a_{u_j}(x_i) \tag{4.3}$$

Then, the sparsity constraint of each hidden unit is added to the optimization objective to minimize an error between it and the given sparsity parameter. For training SAE, root mean square propagation is proposed to improve computational efficiency by minimizing an error and the cost.

4.4.3 Classification

The performance of the new representation of features is evaluated on the validation set of data using a classifier. After training the SAE, a deep 2-D CNN is built, which it takes MRI scans as input. The details of training the built CNN are listed in Table 3.1.

For optimizing the deep neural network, back propagation algorithm is often used for training it to minimize cost and speed up convergence rate [17]. In the proposed approach, stochastic gradient decent, one of popularly used back propagation algorithms, is used for CNN training by adjusting its hyper-parameters, including weights, learning rate, and biases.

For comparative analysis, LDA and Logistic regression have been used for testing the proposed approach. Given the whole low-dimensional image as input, the output of classifier is sensitivity, specificity, and accuracy of true prediction of five types of dementia.

- Convolutional Neural Network: After training CNN, the extracted features of each subject in the testing set is fed into the last output layer of CNN that is softmax layer to predict the type of dementia. Then, correct prediction accuracy is computed the network and is diagnosed by the layer of network, logistic regression layer. Then, correct prediction accuracy is then computed for each diagnosis as a means of evaluating the algorithm's performance.

4.5 Experimental Results

The extracted SAE features are classified by using CNN, LDA, and logistic regression classification models so that the classification accuracy of five types of dementia is calculated and compared. Then, classification accuracy using SAE and PCA generated low-dimensional features is compared. For conducting experiments, we selected 734 MRI images collected from 113 subjects in each one of following classes: elderly controls (HC), patients with Frontotemporal lobe dementia (FTD), patients with uncertain dementia (UD), patients with incipient PTP dementia, and patients with dementia of Alzheimer's disease (DAT). In order to evaluate the efficiency of the high-level features, classification methods are applied on these features.The proposed method was trained used a training set of 520 examples. A testing set of 214 examples was used to evaluate the model' s performance. In order to evaluate the efficiency of the high level features, classification methods are applied on these features. TensorFlow is a deep learning framework that is used for implementing this study.

4.5.1 Data Description

Experiments are conducted on MRI images made available as part of Open Access Series of Imaging Studies (OASIS) datasets analyzed in [97] for evaluating the performance of the proposed model.The description of this dataset is shown in Table 3.2, provided in Chapter 3.

4.5.2 Classification Results

To enhance the performance of the proposed model when deployed to make a prediction on a new unseen data, seven-fold cross validation is performed on 734 MRI images. Cross-validation could assist in overcoming any overfitting problem that could occur due to model complexity thereby providing a good estimation of the accuracy of proposed CAD method in predicting multi-type dementia. After reducing the dimensionality of the dataset, seven random datasets are produced to repeat training and testing of the model in which 60% of MRI images are assigned to the training set while 40% of images are used for testing the algorithm. Moreover, various state-of-the-art techniques including LDA, logistic regression as well as CNN in [80] are adopted on the same data to evaluate the robustness of the proposed method. For comparative analysis of the proposed method and some state-of-the-art models, various performance measurements are used including classification accuracy, sensitivity, and specificity.

The first experiment is conducted in this chapter to test the hypothesis that training the proposed model with features learned by SAE provides higher accuracy of the correct prediction for different types of dementia than training it on the raw pixel intensities. As shown in Table 4.1, maximum testing accuracy reached, sensitivity, and specificity are computed for comparing the performance of the proposed algorithm and the performance of some state-of-the-art methods after training them on the selected dataset. However, the clinical diagnosis is used as the target during training and testing procedures. As a result, Table 4.1 indicates that the accuracy of our proposed model that combines SAE and CNN and then classifies different types of dementia by the output layer outperforms classification accuracy of other state-of-the-art methods. Table 4.2 provides the confusion matrix of four-way classification with features extracted from MRI images.

Table 4.1: Accuracy of various classifiers using SAE for dimensionality reduction in the prediction of dementia.

Machine learning model	Testing Accuracy	Sensitivity	Specificity
LDA	71.8%	**80.86%**	20.25%
Logistic regression	76.2%	35.72%	65.0%
Gupta et al. [80]	70.04%	25.29%	64.15%
proposed method	**80.1%**	63.82%	**75.49%**

Table 4.2: Confusion Matrix of the proposed method.

Actual	No dementia	Uncertain dementia	HC	DAT
DAT	7	2	27	1
Uncertain dementia	2	1	6	0
HC	25	15	203	2
FTD	0	0	3	0

The second experiment is conducted to test the hypothesis that high-level feature representation with SAE would get better classification performance results than features learned by PCA. By comparing the SAE features to the PCA generated features, we report that SAE outperforms PCA in obtaining higher accuracy of diagnosing dementia. In this experiment, PCA decomposes the selected dataset in three orthogonal components that have the maximum amount of the variance, so that principal axes in feature space represent the directions of maximum variance in the data; The overall accuracies are illustrated in Table 4.3. Therefore, we could say that SAE can preserve a lot of details for MRI images after compressing them in a comparison to PCA.

Table 4.3: The comparison between the accuracies obtained by autoencoder and PCA in a combination with each of Machine Learning techniques.

Machine learning model	Autoencoder	Principal Component Analysis
LDA	71.8%	63.2%
Logistic regression	**76.2%**	**68%**

4.6 Summary

This chapter proposed a new method that included the following steps: normalization, image preprocessing and segmentation, feature abstraction as dimensionality reduction, and classification. This method applied the SAE for learning a low-dimensional representation of input data in order to improve the classification task. The proposed approach and some state-of-the-art methods were tested with OASIS dataset. The experiment indicates that our method, combining the SAE with the CNN performs better than other compared state-of-the-art methods. The experimental results showed that the proposed approach yields higher classification accuracy compared to that using the PCA by 8 percent of classification accuracy. The experimental results showed that use of features learned by SAE in early diagnosis of dementia provides better classification performance than the use of raw image pixel intensities for diagnosing dementia. Specifically, the proposed model performs better than PCA in reducing the dimensionality of input data. As future work, the number of sparse autoencoders for reducing dimensionality could be increased to boost the performance of the model, thereby enhancing the accuracy of prediction of multi-type of dementia.

The upcoming chapter proposes a Stacked sparse autoencoder (SSAEs)-based learning approach to learn high-level features from a huge MRI neuroimage dataset.The features extracted by applying the SSAEs are taken as input of the CNN. Finally, the learned features by the CNN can be used for the classification of 4000 MRI neuroimages to improve the early diagnosis of multi-type dementia.

Chapter 5

Context-aware CAD using Multi-Classifier Fusion

5.1 Introduction

Early diagnosis of dementia and its type can strongly influence its prognosis [51], however, accurate identification of various types of dementia is a challenging task. Types of dementia as well as other brain disorders share a set of neurotic pathology. For example, Lewy Bodies dementia has many of the clinical and pathological characteristics of the dementia that occurs during the course of Parkinson's disease. This is further compounded by the need for early diagnosis in which medical brain scans can show limited biomarkers. Thus, neuroimages and medical brain scans may not be sufficient for accurate identification of dementia. Accordingly, this work aims to utilize additional medical and context information to enhance the accuracy of identification of the various types of dementia.

Accurate identification of various type of dementia can be achieved by the use of the structural and functional biomarkers extracted from neuroimages coupled with the contextual data that interprets images and describes patients' status. The augmentation of neuroimaging analysis with contextual information is challenging due to the discrepancies and irregularities of the various forms of data. Multi-classifier fusion has proven to be a powerful tool to improve the classification accuracy when data is characterized by high discrepancies and irregularities. Using multiple individual classifiers can provide a joint decision by combining their individual output in order to derive the final result. Accordingly, we hypothesize that augmenting the features extracted from neuroimages with contextual and medical data of the patients through multi-classifier decision fusion would enhance the classification accuracy.

This chapter presents a novel context-aware CAD for early identification of multiple type of dementia by augmenting the medical imaging data with contextual patients' data through the decision fusion of multiple classifiers. This novel approach can help in building a robust model that offers a high degree of diagnostic accuracy regarding the identification of multiple types of dementia. The rest of the chapter is organized as follows: Section 5.2 briefly describes a stacked sparse autoencoder and reviews the related work on multi-classifier fusion. Section 5.3 presents the framework of a context aware approach proposed to automatically detect several types of dementia. Section 5.4 reports the significant experimental results, and Section 5.5 summarizes the chapter.

5.2 Background and Related Work

5.2.1 Stacked Sparse Autoencoders (SSAEs)

Recent reports have proposed the use of stacked sparse autoencoders (SSAEs) in order to reconstruct original input data to fit in a smaller representation and to enable the features of a dataset to be learned in an unsupervised manner. Both tasks are accomplished through the minimization of the reconstruction error between the input data at the encoding layer and the data reconstruction at the decoding layer. The goal of these tasks is to improve AD detection [87]-[88].

SSAEs are deep neural networks made up of multiple layers of basic sparse autoencoders in which the output of each layer is wired to the input of the successive layer. SSAEs can be pre-trained using a greedy layerwise approach that involves training each layer in turn. The first SSAE layer tends to learn the first-order features of the raw input. The second layer is then focused on learning second-order features that correspond to patterns in the appearance of the first-order features. Higher SSAE layers are generally directed toward learning even higher-order features.

5.2.2 Multi-Classifier Fusion

An ensemble of classifiers is a set of classifiers that are trained over features extracted from a dataset. Their individual output results are then combined using any of the fusion techniques to obtain a final classification score [89]. Classifier fusion has recently become viewed as one of the most promising methods in the field of pattern recognition. The design of multi-classifier fusion involves two phases: the design of multiple classifiers and the design of a combination technique (fuser).

A variety of multi-Classifier fusion methods have been recently developed, suggested as alternative approaches leading to a potential enhancement in classification performance. These methods include weighted averaging, fuzzy integral methods, ranking methods, Dempter-Shafer method, and weighted voting [105]. These existent methods for the combination of multiple classifiers can be classified into three frameworks, i.e. linear opinion pools, evidential reasoning, and logarithmic pools.

In recent years, few multi-classifier fusion algorithms have been proposed for early prediction of multiple stages of Alzheimer' s disease severity [90,91,103,104]. In 2013, Liu et al. proposed a novel multi-classifier fusion approach to differentiate between patients with AD and healthy people with the weighted voting strategy. Then, the outputs of all high-level classifiers were combined for making a final classification. Their results revealed that classification was enhanced by combining the output of multiple classifiers, which provided an accuracy level of 92.0%, and 85.3% for classifications of AD versus NC and MCI versus NC, respectively [90]. In [91,103], new contributions were introduced in order to improve classification performance with respect to multi-stages of AD. Both works proposed multiple classifiers fusion approaches with a majority-voting rule in order to obtain the final prediction. After extracting features that capture AD biomarkers, small set of these features were selected for classification tasks, including volume of GM, WM, and CSF and size of hippocampus. Their results showed that building multiple classifiers outperforms the use of a single classifier with respect to accurately predicting AD.

5.3 Methodological Approach

This research work aims to boost the accuracy of early prediction of dementia through augmenting the medical imaging data with contextual patients' data. In contrast to the state-of-the-art approaches described above, the proposed CAD approach differentiates among various types of dementia, whereas most of the reported studies have been focused on differentiating among the varied stages of AD severity. While all aforementioned studies focused on evaluating features extracted from MR neuroimages, contextual information of dementia, that can be inferred from the metadata associated with neuroimages, are rarely considered or discussed in dementia classification literature. For example, types of dementia can be more confidently labeled by considering scores of clinical diagnosis tests like clinical dementia rate and mini-mental examination tests. This work proposes a novel CAD approach, which integrates features extracted from MR neuroimages as well as contextual information instead of only using the extracted features for the classification purpose. This objective is accomplished through the decision fusion of multiple classifiers. Figure 5.1 illustrates the proposed hierarchical structure of the context-aware multi-classifier fusion framework.

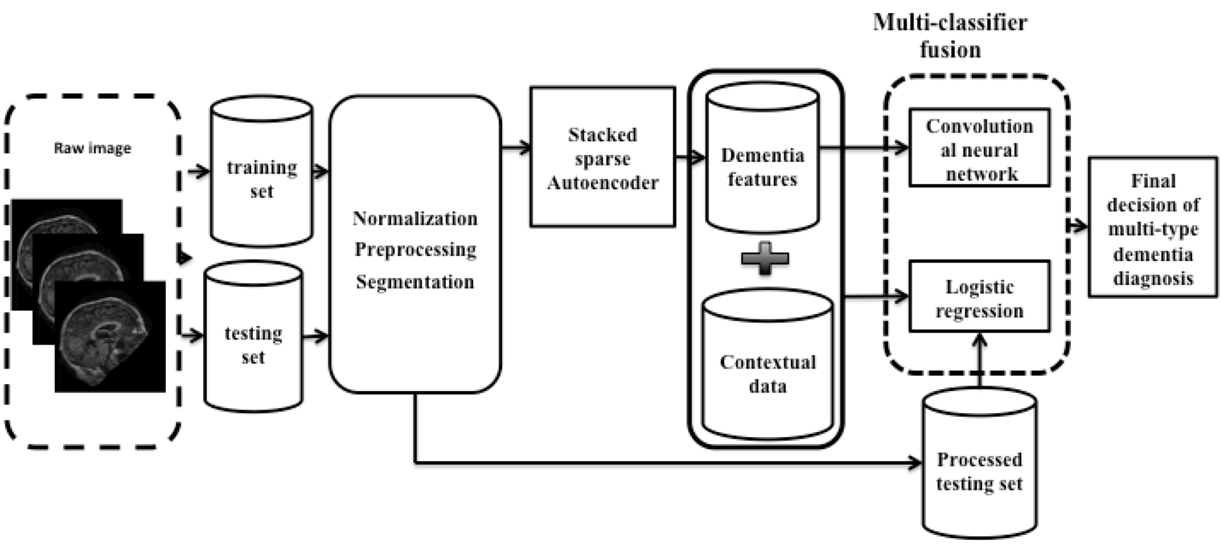

Figure 5.1: Block diagram of the proposed context-aware method for improving the early detection of multiple types of dementia and for predicting classification accuracy.

As shown in Figure 5.1, the proposed framework initially preprocesses, and segments magnetic resonance imaging (MRI) scans to normalize and remove all portions surrounding the brain. An SSAE is pre-trained with regions of interest, hippocampous and gray matter tissue, to extract more sophisticated features that capture multiple types of dementia biomarkers, and reduce the high dimensionality of the images. Subsequently, the whole MR images are classified based on features learned by the SSAE along with context information using two classifiers, the convolutional neural network (CNN) and logistic regression. The CNN output and the logistic regression output are then combined by the linear opinion pool to obtain the final classification decision.

Image preprocessing and segmentation are detailed in Section 5.3.1, and Section 5.3.2 explains feature extraction and dimensionality reduction, followed by a description of the classification fusion technique in Section 5.3.3.

5.3.1 Image Preprocessing and Segmentation

An image preprocessing and segmentation as described in chapter 4 will help to remove the skull portion surrounding the brain tissues and to identify and measure the regions of interest, i.e. hippocampus, and gray and white matter tissues, required for dementia classification. For preprocessing and segmenting medical neuroimages, FreeSurfer software is used. Using FreeSurfer is discussed in detail in Chapter 4.

The rationale of this study is the enhancement of classification performance for early diagnosis of multiple types of dementia by advancing the CAD method, presented in Chapter

4. In addition to using MR neuroimages in classification, some data can be inferred from metadata associated with images to improve classification performance. Such data serves as contextual information that can be a complementary to the information extracted from images for training and classification. This context information is converted into features that can be understood by a classifier.

5.3.2 Feature Extraction and Dimensionality Reduction

Feature extraction from medical images is critical significance of increasing the accuracy of the CAD method in early prediction of multiple types of dementia. When the used features for training deep-learning techniques are insufficient or imprecise, this often leads to poor classification. For example, a sparse autoencoder was pre-trained on MR neuroimaging data to learn a latent or compressed representation in the earlier chapter. From the experimental results, it was demonstrated that the sensitivity of SAE-based model is worse than specificity of one of comparing state-of-the-art studies. This may happen due to the simple shallow structural characteristic of SAE, which restricts the extraction of high-level features. To modify the extracted features and build more robust learning method, this chapter proposes to use the Stacked Sparse Autoencoder (SSAE) for extracting high-level features and reducing the high dimensionality of large imaging dataset.

The SSAE is composed of two layers that represent the sparse autoencoder (SAE).

Figure 5.2 illustrates the SSAE model with two autoencoders stacked hierarchically. The SSAE is applied in order to learn the latent features of the raw 2D images that have been input. The output of each SAE is wired to the input of the following layer. Each layer consists of three input/output mapping functions: an encoding function *f(x)*, a feature for mapping the data to compressed data, and a distance function that computes the amount of information lost between the compressed representation of the data and the decompressed representation. A final feature is a decoding function *g(x)* that maps the compressed data to low-dimensional data, which is approximated to the original data.

In the input layer, the raw 3D images, $x \in R^{n-input}$, are converted to pixel intensity values so that each image is represented as a column vector of pixel intensity whose size is 176 2×1 . The input layer contains n-input=$176 \times 176 = 30976$ input units. Input x is fed into first trained SAE with 5000 hidden units, which is then encoded so that it learns primary features, h_1, for each of input x and reduce its dimensionality by adjusting the weight, W_1, as given in Equation 5.1.

$$h_1 = f(Wx + b) = W_1 x + b_1 \tag{5.1}$$

where $W \in R^{u \times n_i nput}$and $b \in R^u$ represent the matrix of the weight and the bias vector of the encoder function, respectively. with respect to the activation function, this study was based on consideration of a logistic sigmoid function for $f(a) = 1/(1 + exp(-a))$, which is the one most widely used in the field of pattern recognition or ML. These primary features are then used as input for the second SAE with 1000 hidden units, so that it can learn secondary features, h_2, for each of primary features h_1 by adjusting the weight W_2. The secondary features are given by Equation 5.2.

$$h_2 = f(W_1 x^* + b_1) = W_2 x + b_2 \tag{5.2}$$

where $W_1 \in R^{u \times n_i nput}$and $b_1 \in R^u$ represent the trained matrix of the weight and the bias vector of the encoder function, respectively. In the final stage, the extracted features are treated as input to the CNN, which is trained to map secondary features to more sophisticated features so that h_2 is decoded to obtain x*, which is a low-dimensional representation of input x. In this way, the extracted features are classified as identifying one of the several types of dementia. This greedy layer-wise learning is called "pre-training".

5.3.3 Classification

Chapters 3 and 4 have proposed features-based deep-learning approaches, so that these approaches are trained directly on features extracted from MR images for identification of dementia. As a result, it has been shown that predicting the type of dementia for some input images are not correct. This may result from sharing the same clinical and patho-

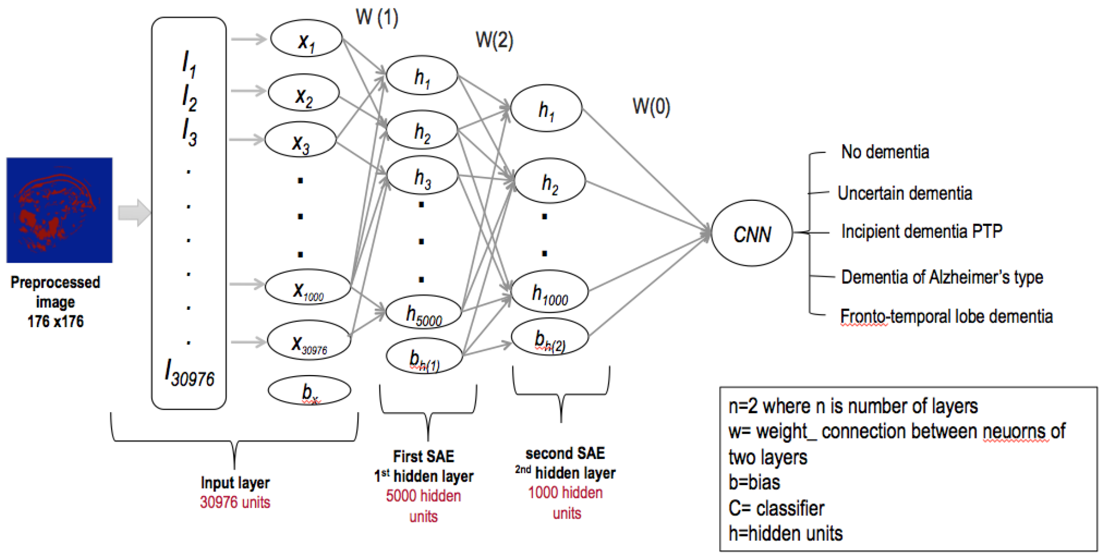

Figure 5.2: A deep architecture of our stacked auto-encoder.

logical characteristics among various types. So, the extracted features from neuroimages and medical brain scans may be not adequate for accurate identification of dementia. To address this shortcoming, this chapter proposes to combine features learned from images and the complementary context information in order to enhance classification performance.

These two kinds of data have been used by the set of base classifiers for recognizing type of dementia for each image and then producing the probability vector for each image. two probability vectors resulted from the CNN and logistic regression are fused by the linear pooling rule into a single probability vector.

- **Convolutional Neural Network (CNN):** The architecture of CNN for feature extraction and classification, discussed in chapter 3 is used in this chapter. For training

CNN in this study, many experiments are conducted to optimize its hyperparameters, including learning rate, number of filters, and filter size.After tuning and optimizing CNN' s parameters, a softmax function is applied in order to convert raw scores into normalized probabilities that represent possible categories of types of dementia.

- **Logistic regression:** The high-level feature vector learned by SSAE are also treated as input to the logistic regression for classifying into several types of dementia. In addition to these features, some additional data has been inferred from metadata associated with images that may potentially enhance classification accuracy. Such data serves as contextual information that can be a complementary to the information extracted from images for classification. This context information is converted into features that can be understood by a classifier.

  Given dimension-reduced training samples *x'*,the feature vector reconstructed by SSAE as well as contextual data as the inputs, the logistic regression is trained for differentiating multiple types of dementia. For multi-class classification, logistic regression model uses the softmax function, a probabilistic function, which calculates the probabilities *p(y/x)* for the given types of dementia y from a combination of the extracted features from input *x'*. This function chooses the output label y that maximizes the probability as given by equation 5.3.

$$y* = argmax_y \, p(y/x) \tag{5.3}$$

The outputs of two classifiers are combined based on the weighted averaging rule to obtain the final decision score using equation (6). For a sample i in a dataset D with five classes, we assume that there are *P(P>1)* feature vectors, which can be extracted from sample i called $x_1, .., x_P$. Accordingly, we employ two classifiers C_1 and C_2, to learn the classification task using features extracted from raw data in D, respectively, in which the input of classifier C_j is the feature vector x_{p_j} $(1 \leq p_j \leq P)$. The output of each base classifier C_j is as follows, that have been trained over the same training set.

$$p(C_j, x) \leq 0, \sum_{j=1}^{5} p(C_j, x) = 1 \tag{5.4}$$

where $p_j(x)$ denotes the probability that sample i belongs to label L_k recognized by the classifier C_j with the input vector x_pj , x denotes a feature vector of sample i with its form of representation being a vector.

This work proposes the set of base classifiers to recognize type of dementia for each image and fuses their outputs using the linear opinion pooling rule. This rule involves taking a weighed linear average of classification outputs combined from various classifiers. In this research, linear opinion pool decision fusion method works on fusing probability vectors resulted from CNN and logistic regression into a single probability vector by using equation (7).

$$L(p_1, .., p_n) = \sum_{i=1}^{n}(p_i w_i) \tag{5.5}$$

Where outputs of classifiers are represented by probability vectors p1,..,pn. $p_i = (p_{i,1}, .., p_{i,c})$ represents the probability vector of each classifier in which c referes to a number of target labels. L is the pooling operator, and w_i denotes the weight associated with classifier i's probability vector so that

$$\sum_{i=1}^{n}(w_i = 1), 0 \leq w_i \leq 1 \tag{5.6}$$

5.4 Experimental Results

To measure the performance of our proposed approach, we have used two classifiers that perform linear pooling decision fusion. Seven-fold cross validation strategy is applied for the evaluation of our approach. The dataset was randomly partitioned into seven folds, and six of the seven folds were then used for training, with the remaining one for testing. This entire process was repeated seven times in order to obtain an unbiased evaluation. The purpose of choosing cross validation is to optimize the hyperparameters of the classifiers. Moreover, cross-validation is used to ensure the best generalization ability of the proposed model on used independent dataset

The effectiveness of the proposed model based on the measurement of metrics such as sensitivity, specificity, precision, f1-score, and the accuracy of the diagnosis for the early detection of dementia. These evaluation measures were implemented in the Python scripting language (version 2.7.11) using the scikit-learn library. The proposed deep neural network model was built using TensorFlow and trained on Google Cloud Platform. In this study, MRI structural data and contextual data interpreting images are used for training the proposed CAD model to categorize types of dementia.

5.4.1 Dataset

Similar to works proposed in the previous chapters, experiments in this chapter are conducted on the OASIS dataset. This chapter uses the entire dataset that consists of 4000 T1-weighted MRI scans, collected from 417 subjects and divides it into three subsets, including training dataset, validation dataset, and testing dataset. To increase classification performance, some additional data can be used. This data includes clinical examination, and demographic and anatomic values. Table 5.1 lists details of the selected dataset. The following data, taken from metadata associated with images in the selected dataset, is used for conducting our experiments:

- Demographical data: Considerable demographic data has been used for conducting our experiments and enhancing the performance of the proposed model including an age of subjects, their gender, their educational status, and their handeness.

- Clinical dementia rate (CDR): The CDR is a five-point scale used to assess the cognitive performance of each patient in different domains and so quantify the severity of dementia. The scale ranges from zero to three, and a value greater than zero indicates that subjects diagnosed with dementia, whereas all healthy control subjects had CDR of 0.

- Mini-Mental state examination (MMSE): The MMSE is a test that can be used to evaluate mental status by answering questions about different cognitive areas. However, the score of MMSE, ranges between zero and 30, so any score greater than or equal to 24 points (out of 30) indicates subjects have normal cognition.

- A number of whole brain volume (nWBV): nWBV denotes the percent of all voxels within the brain mask that are classified as gray or white matter' s tissue.

Table 5.1: Contextual data of patients in OASIS dataset.

Attribute	Description	Values	Num. of subjects
Age	Age in years	18-25 26-45 46-65 66-96	120 48 71 178
Sex	Gender of patient	Male Female	160 257
CDR	Clinical dementia rate	0 0.5 1	89 16 8
MMSE	Mini-mental state examination	18-23=demented 24-30=NC	
nWBV	Normal of whole brain volume	0.644-0.797=demented 0.645-0.893=NC	
Sequence	MR image sequence	3D T1-weighted images	

5.4.2 Performance Evaluation

To evaluate the effectiveness of the proposed approach in predicting multiple types of dementia, a number of metrics are used, including classification accuracy, sensitivity, specificity, precision, and f1-score. In addition to these metrics, precision-recall curve and micro-averaging are also used for evaluating the model' s performance in predicting each class of dementia in the dataset. Furthermore, the pre-mentioned metrics are used for comparing the performance of the proposed model with the performance of that in earlier chapters.

5.4.2.1 Performance Evaluation of the Proposed Approach:

The results clarify how incorporating neuroimages with the contextual data affects classification performance as it will be discussed. Table 5.2 shows that the proposed model yields an overall classification accuracy of 95% of five types of dementia with a specificity of 93%and a sensitivity of 89%. From these results, it can be confirmed the effectiveness of the proposed approach and its merit in increasing classification accuracy by fusing of multiple classifiers that use the extracted features and contextual data for training and classification. Figure 5.3 shows accuracy during training and after testing. The results show that the convergence between training accuracy and testing accuracy curves is increased until 200k epochs and starts to decrease.

Table 5.2: The results of classification performance using a combination of ML classifiers.

Performance metrics	Percentage (%)
Training Accuracy	95
Testing accuracy	90
Sensitivity	93
Specificity	88
Precision	94
F1-score	93

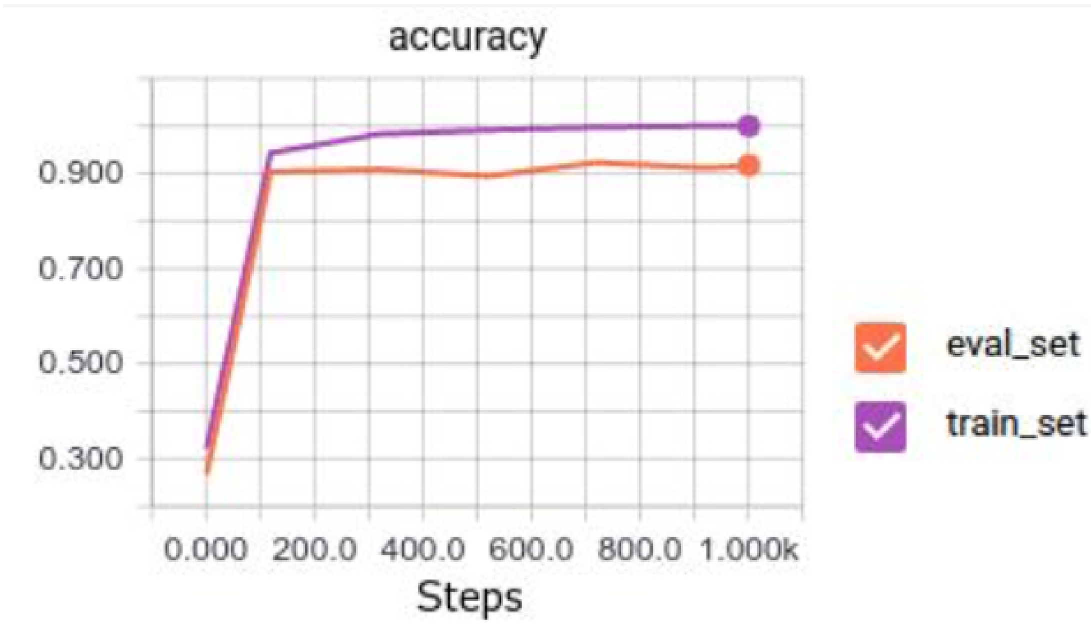

Figure 5.3: Training and testing accuracy of performance of proposed model.

5.4.2.2 Performance Comparison of the Proposed Approach with that of Pervious Chapters:

The second experiment is performed to test the effectiveness of proposed approach which augmenting data with patient's context information through fusing multiple classifiers. Specifically, we compare the performance of the proposed multi-classifier fusion method with those of other three possible classification methods presented in previous chapters. The classification results and their comparisons with respect to different methods are summarized in Table 5.3. Regarding augmenting data with contextual data, it has been observed that the ensemble of classifiers is a good option for the overall classification between dementia subjects and healthy controls due to its superior accuracy rate compared to its base classifiers. We have been able to achieve higher sensitivity, specificity, and accuracy rates as compared to other competing approaches [102]-[106] (see Table 5.3) for identification of several forms of dementia.

Table 5.3: A comparison of the proposed method with three methods presented in the previous chapters in terms of classification accuracy, specificity, and sensitivity.

Classification methods	Accuracy	Sensitivity	Specificity
The proposed CNN-based CAD	75%	63%	46%
The proposed SAE+CNN CAD	80.1%	63.82%	75.49%
The proposed SSAE+CNN CAD	83.69%	79.32%	77%
The fusion-based approach	**95.0**%	**93**%	**88**%

5.4.2.2.1 Discussion

The dataset used to conduct the experiments is imbalanced. The dataset has five labels, representing types of dementia, to be classified. Due to splitting dataset into training and testing dataset randomly, training dataset contains a limited number of examples in different classes of dementia. The highly imbalanced dataset leads to overfitting, thereby negatively affecting classification performance.To boost the classification performance, some ways are performed to compact imbalanced training dataset before providing the data as input to the proposed method. This improvement is accomplished by tuning the hyper-parameters of the CNN and the SSAE, modifying the proposed classification algorithm, and augmenting and resampling data. This section presents performance results of the proposed model with imbalanced data as well as the results of the proposed model after handling data.

5.4.2.2.2 Experimental Results with an Imbalanced Dataset without Data Improvement:

Due to working with very imbalanced classes in the selected dataset, some measurements instead of accuracy have been designed to measure the success of prediction of multiple types of dementia. These metrics include the precision-recall curves, confusion matrix, and micro-averaging. The Precision-Recall curves for all labels of dementia are plotted on the same graph, and then the area under the curve (AUC) is calculated for all classes (see Figure 5.4). As illustrated in this figure, the proposed algorithm is biased towards the majority class, where the precision of the diagnosis of normal controls is 83%, whereas the precision of classification of other remaining types of dementia is under 8%.

In addition to plotting precision-recall curves, the micro-averaging is also calculated in order to evaluate the performance of the proposed approach in early prediction of five types of dementia by summing all true positives, true positives, and false negatives for each class label, and then applying them to get classification performance. Figure 5.5 illustrates an area under the micro-averaging curve, which represents the average precision of the dementia diagnosis. AUC is 62 % which has been computed from prediction scores. The

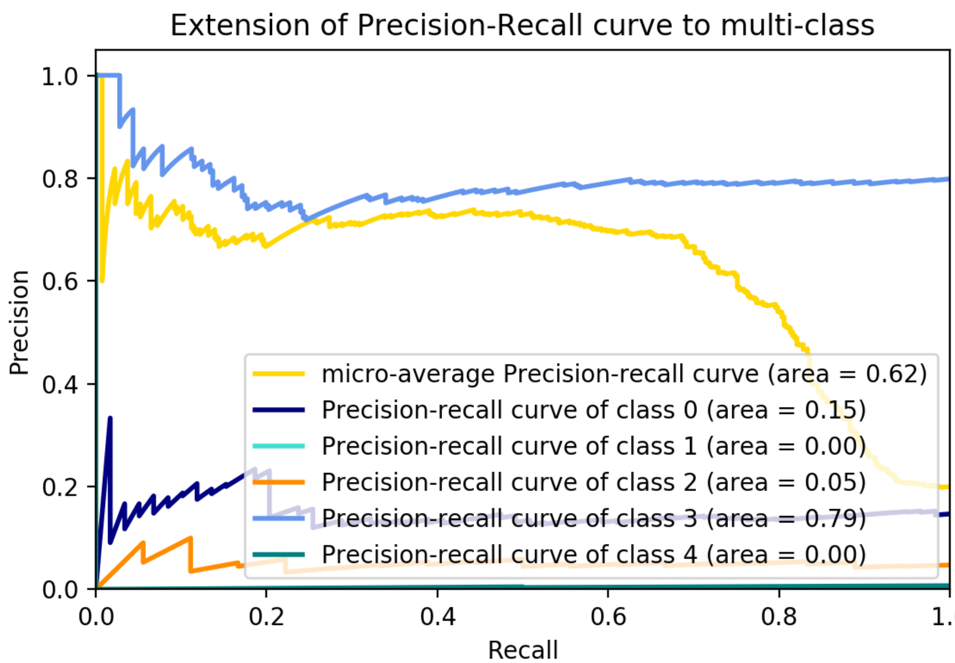

Figure 5.4: The precision-recall curves for all classes of dementia.

results in this figure show that AUC of the last class, uncertain dementia, in the dataset is nan that means no any instance with this class is correctly predicted.

In addition, Figure 5.6 shows the confusion matrix of the five-way classification with augmenting the features extracted from the MRI images and the contextual data. The results presented in this confusion matrix show that the CAD approach yields poor predictive accuracy for the minority classes, including AD, FTD, uncertain dementia, and incipient dementia PTP and it tends to classify most new samples in the majority class which is normal controls. However, these unsatisfied results are happened due to the use

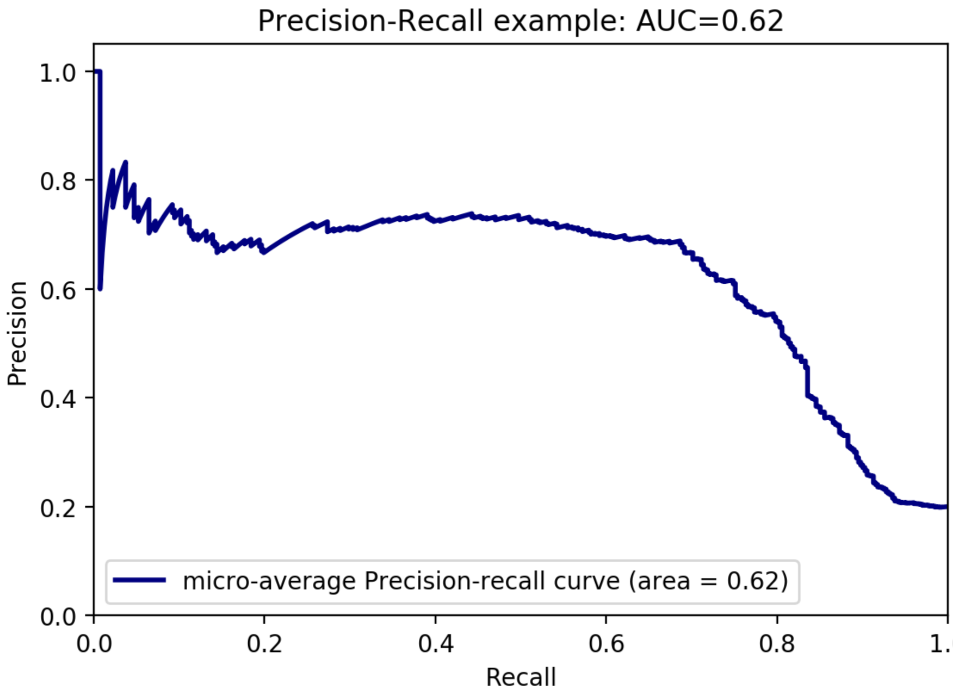

Figure 5.5: An average precision score, micro-averaged over all classes.

of highly imbalanced dataset.

5.4.2.2.3 Experimental Results after Handling Imbalanced Dataset:

In Section 5.4.2.2.2, it has been notable that classification is biased in favor of the majority class. This problem is attenuated by manipulating MR neuroimages before training the proposed model whereby a variety of augmented images for each of minority classes are generated. The proposed CAD approach tested patients clinically diagnosed as having dementia due to AD, FTD, uncertain dementia, or incipient dementia PTP. It correctly

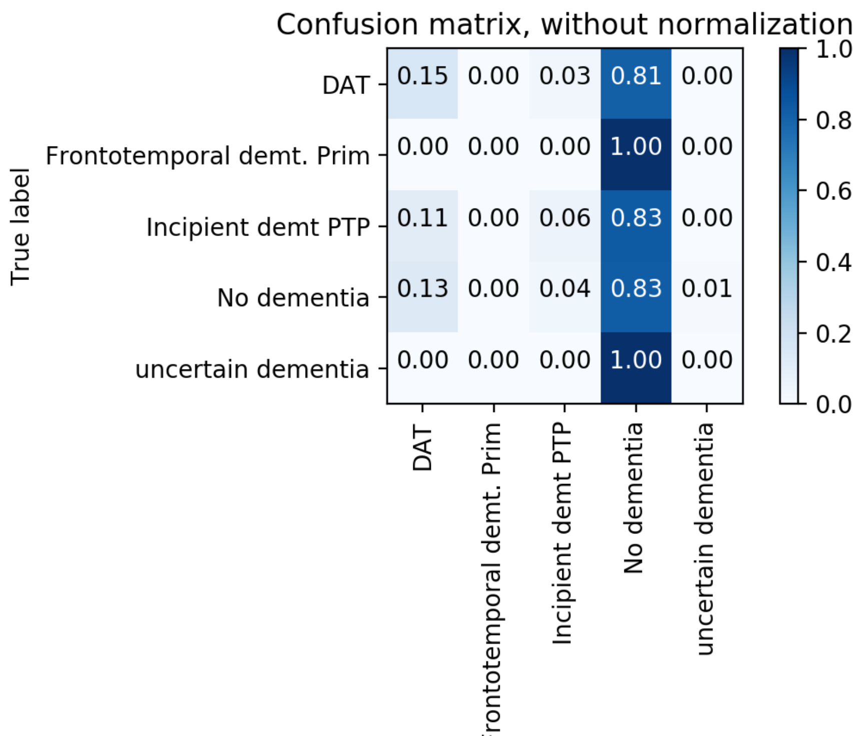

Figure 5.6: Confusion matrix of identification of five types of dementia.

classified 400 out of 500 samples. Figure 5.7 shows the confusion matrix of the five-way classification after up-sampling the minority classes. In this figure, the five diagonal cells show the percentage of correct classifications by the trained CAD model. For example, 51 samples are correctly classified as patients having AD. This corresponds to 86% of all 500 samples. Similarly, 51 cases are correctly classified as patients with FTD. This corresponds to 100% of all samples. Moreover, 316 out of 335 normal controls are correctly classified so that it corresponds to 94%. Also, all patients with incipient dementia PTP and patients diagnosed with uncertain dementia have been classified correctly. This corresponds to 100%.

A nine out of 500 samples are incorrectly classified as having AD, and this corresponds to 0.15. Similarly, 19 of patients with different dementia types are incorrectly classified as normal controls and this corresponds to 5% of all data. Out of 325 no dementia predictions, 94% are correctly predicted as normal controls and 15% are predicted as AD. Similarly, out of 65 AD predictions, 86% are correct and 4% are wrong predictions. Overall, 94% of the predictions are correct and 6% are wrong classifications.

In addition to group classification, Figure 5.8 and Figure 5.9 illustrate the multi-class precision-recall curves and the area under the micro-averaged curve (AUC) to visualize the performance of the proposed model in the prediction of multiple types of dementia. The results in Figure 5.9 clarify that the AUC for precision is significantly higher than that

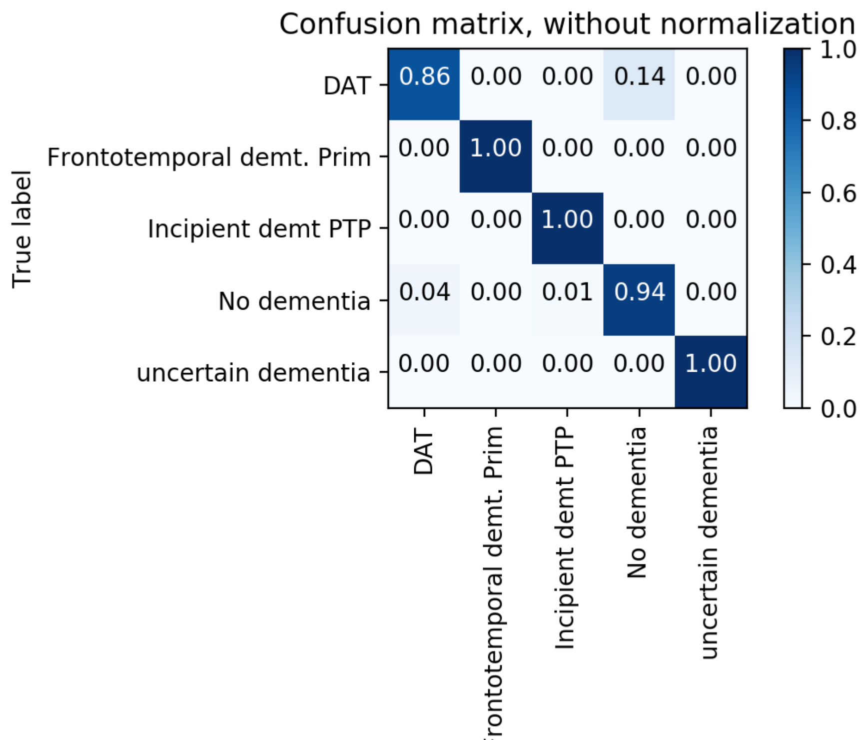

Figure 5.7: Confusion matrix of identification of five types of dementia after manipulating imbalanced data.

computed using the imbalanced dataset in the previous section. It has been shown that average precision is up to above 93%. This score of micro-averaging shows that the proposed CAD model is returning accurate results (high precision), as well as returning a majority of all positive results (high recall). Moreover, Table 5.4 summarizes some classification statistics in the prediction of several types of dementia, including the precision, recall, and specificity along with the F1 score for each label after handling its bias. The results displayed in that table demonstrate that proposed classification method can achieve high classification performance due to manipulating high dimensional imbalanced data.

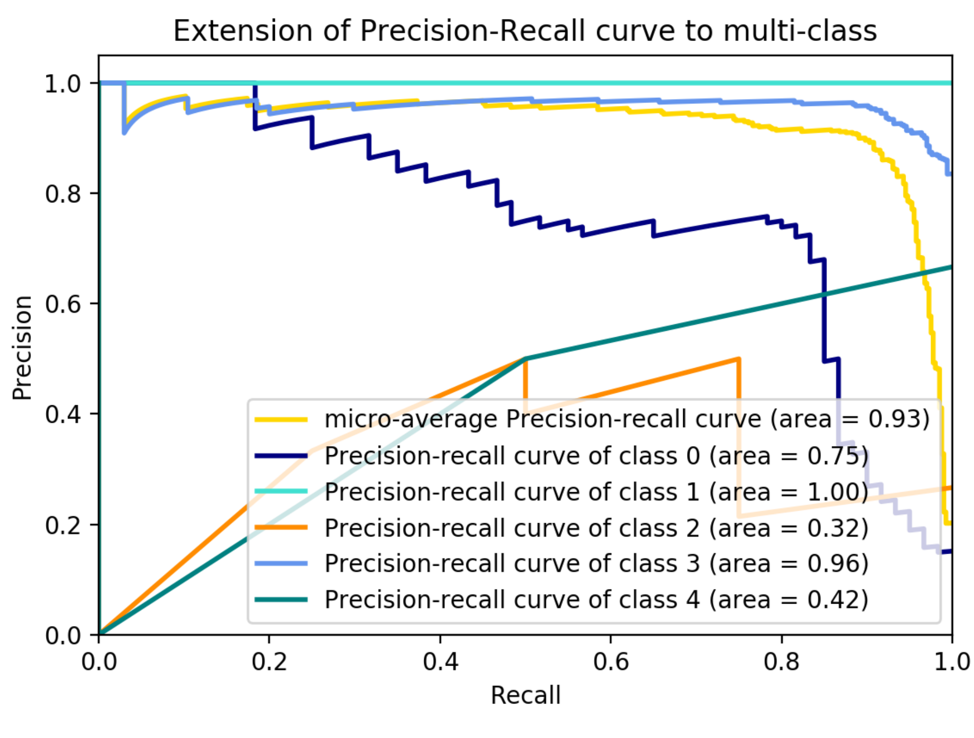

Figure 5.8: The precision-recall curves for all classes of dementia.

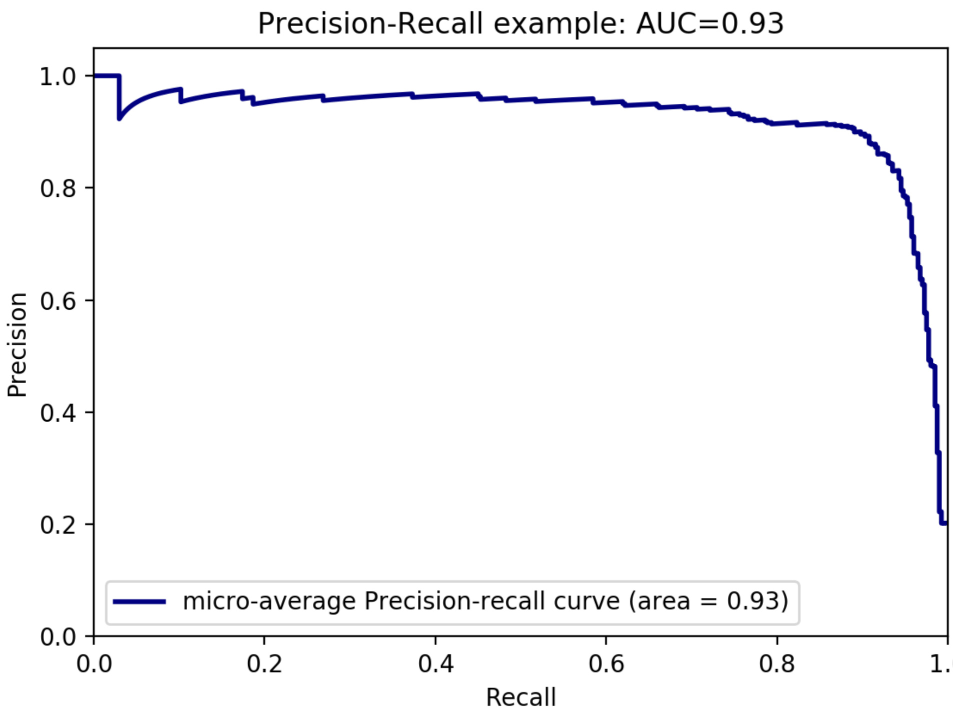

Figure 5.9: An average precision score, micro-averaged over all classes after balancing dataset.

Table 5.4: A classification report displaying percentages for each of precision, recall, specificity, and f1-score of each class.

Class label	Precision	Recall	Specificity	F1-score
Alzheimer's disease	78%	86%	96%	82%
Fronto-temporal dmentia	100%	100%	100%	100%
Incipient dementia PTP	70%	100%	99%	77%
No dementia	97%	94%	87%	96%
Uncertain dementia	67%	100%	100%	80%
avg/total	94%	93%	88%	93%

5.5 Summary

This chapter has described the investigation of a new context-aware CAD approach for early differentiation of multiple types of dementia based on MRI scans and the contextual data. An ensemble of classifier-based automated classification algorithms has been proposed for gradually combining individual decisions into a unified model, in this case, based on linear pooling decision fusion rule, in order to enhance classification performance. The specific process involves first the building of a CNN for transforming the features learned by an SSAE into more compact high-level features through unsupervised learning. A logistic regression classifier and a neural network are then generated, each evaluating the high-level features of different brain regions. In the final stage, the classification outputs of both classifiers are combined to obtain the final classification decision. The proposed method has been evaluated using MR neuroimages of 417 subjects from the OASIS dataset.

The experimental results revealed that the proposed learning algorithm achieved an accuracy level of 95 %, a sensitivity level of 93 %, a f1-score of 93 %, a precision of 94%, and a specificity level of 88 % for multi-type dementia classification, thus demonstrating very promising classification performance.The correct classification is 93% for all groups, 86 % for AD, 94% for NC, 100% for incipient dementia PTP,100% for uncertain dementia, and 100% for FTD.

Chapter 6

Conclusion and Future Work

6.1 Summary and Contribution

This work has proposed a novel deep learning based computer-aided diagnosis (CAD) approach for early identification of multiple types of dementia, whereby features extracted from neuroimages and patients' contextual data are incorporated and utilized to decision process. The architecture of the proposed approach consists of three stages, including image preprocessing and segmentation, feature extraction, and classification. After segmenting images and obtaining the interest regions for detecting dementia, the proposed approach trained Stacked Sparse Autoencoders (SSAEs) consisting of two layers to learn an improved feature space, including the hippocampus' s size, and gray and white matter' s volumes. Each layer of SSAEs learned a combination of features from features produced by the previous layer. Then, the learned features as well as context information were then classified by a combination of classifiers to enhance classification accuracy of multiple types of dementia.

This thesis developed a novel deep learning-based CAD approach on several stages to assist and help practitioners in differentiating among several types of dementia in early stages using MRI scans. Chapter 3 presented a convolutional neural network- based approach for early prediction of multi-type dementia. As a successful deep model applied in dementia diagnosis, the results of Chapter 3 showed that the Convolutional neural network in classification has demonstrated superior performance to the state-of-the-art approaches whether in classification accuracy or sensitivity while the specificity of prediction achieved using the proposed method is not better than the specificity measured by other compared methods. The weakness of specificity can be a result of training a very deep CNN with a small dataset.

In chapter 4, this research work sought to make the proposed model, presented in Chapter 4 more robust and faster through reducing high dimensionality and extracting features from MRI images before feeding them into the CNN for classification. However, this paper suggested computer aided approach for differentiating the different forms of dementia in its early stages by training low representation of MRI images obtained by SAE. This CAD method combines autoencoder that learns high level features with ML techniques that classifies data into one of five dementia classes. The accuracy of this proposed approach was predicted on large patient population. The experiment indicates that our proposed model, merging SAE with CNN, achieves better accuracy than both linear support vector machine and logistic regression. Also, a comparison among the performance of the proposed model and the performance of PCA was conducted. It has been noticed that the results of proposed model outperforms the results of PCA.

Chapter 5 advanced the CAD approach, presented in earlier chapter to enhance classification performance. This was accomplished by increasing training data with patient's contextual information through multi-classifier fusion. The results indicate that the fusion of multiple classifiers achieves better performance results in prediction of multi-type dementia in early stages than building a single classifier for achieving the same goal. The proposed approach has best sensitivity and specificity values as compared to that of existing approaches. The proposed approach has an overall accuracy value comparable to existing approaches despite the fact that we have used smaller feature set. the early identification of multiple types of dementia. While there is no cure for dementia, early effective diagnosis of dementia is essential in managing it. Early diagnosis plays an important role in identifying the right treatment, preventing or slowing down cognitive ability deterioration, and in seeking the appropriate support and planning for the future [28]. However, the early diagnosis of dementia is a challenging task due to the extremely complexity of dementia characterizations in neuroimaging data. The proposed multi-stage CAD approach

6.2 Concluding Remarks

The results of the hypotheses tests are summarized below:

- **H1:** The use of features extracted by the deep learning techniques in the early diagnosis of dementia provides higher classification performance than the use of raw image pixels' intensity in diagnosing dementia. The results of this study indicate that this hypothesis is tested by conducting an experiment explained in Chapter 3. The experiment was an attempt to investigate that using features learned by Convolutional neural Network (CNN) from MRI images of brain improves classification accuracy comparing to the use of raw image pixels for classification purposes. Using deep learning techniques in the early diagnosis of dementia would perform higher classification performance.

- **H2:** Reducing high dimensionality of neuroimages into high-level feature would yields higher classification performance and accurate identification of various types of dementia biomarkers. This hypothesis was tested by conducting the experiment provided on Chapter 4. The proposed CAD approach is compared with state-of-the-art approaches, and it has shown its superiority with 4.1% better performance.

- **H3:** Augmenting the features extracted from neuroimages with contextual and medical data of the patients through multi-classifier decision fusion would enhance the classification accuracy. Chapter 5 provided the experiment, conducted to test this hypothesis (H3). A multi-classifier- based CNN and logistic regression are fused to enhance classification performance. The results showed that the hypothesis is accepted.The proposed CAD approach has shown high performance in early identification of multiple types of dementia, where its accuracy has reached 95%. The model has compared with models, presented in previous chapters,it has outperformed both methods with 75% and 80.1%, respectively.

Bibliography

[1] Ahmed Abdulkadir, Jessica Peter, Thomas Brox, Stefan Klppel, and Olaf Ronneberger. Voxel-based multiclass classification of AD, MCI, and elderly controls: blind evaluation on an independent test set. In *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pages 8-15, 2014.

[2] Thomas Ban. Dementia: Differential diagnosis.In *International Network for the History of Neuropsychopharmacology*, pages 1- 31, 1991.

[3] 2012 Alzheimer?s disease Facts and Figures. *Alzheimer's & Dementia*, volume. 8, pages 1-67, (2012).

[4] Katja Franke and Christian Gaser. Dementia classification based on brain age estimation. In*Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pages 48- 54, 2014.

[5] 2016 Alzheimer' s disease Facts and Figures (2016) [Online]. Available at `http://www.alz.org/facts/`.

[6] 2015 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, Vol. 11, pages 332- 384, (2015).

[7] Anna Dowrick and Alice Southern, *Dementia 2014: Opportunity for change* in Alzheimer's Society, London, (Sep. 2014).

[8] Bala Mahendra, *Dementia: A survey of the syndrome of dementia*, MTP Press Lancaster; 2 edition, 1984. [Online] Available: Springer.

[9] Anthony Akobeng. Understanding diagnostic tests 1: Sensitivity, specificity and predictive values., *Acta Paediatrica*, Volume. 96, pages. 338- 341, 2007.

[10] ChristianWachinger, Polina Golland, and Martin Reuter. BrainPrint: identifying subjects by their brain. *Proc Intl Conf Med Image Compute Comp Ass Intervent.* Lecture Notes in Computer, volume. 8675, pages 59-70, 2014.

[11] Adrien Payan and Giovanni Montana. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks, 2015.

[12] Alessia Sarica, Giuseppe Di Fatta, Garry Smith, Mario Cannataro, and James Saddy. Advanced feature selection in multinomial dementia classification from structural MRI data. *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pages 82- 91, 2014.

[13] Hidenori Yoshida, Seishi Terada, Hajime Honda, Toshie Ata, Naoya Takeda, Yuki Kishimoto, Etsuko Oshima, Takeshi Ishihara, and Shigetoshi Kuroda. Validation of Addenbrooke's cognitive examination for detecting early dementia in a Japanese

population. In *Psychiatry Research*, volume.185, pages 211- 214, 2011.

[14] Esther Bron, Marion Smits, Wiesje Flier, Hugo Vrenken, Frederik Barkhof, Philip Scheltens, Janne Papma, Rebecca Steketee, Carolina Orellana, Rozanna Meijboom, Madalena Pinto, Joana Meireles, Carolina Garrett, Antnio Bastos-Leite, Ahmed Abdulkadir, Olaf Ronneberger, Nicola Amoroso, Roberto Bellotti, David Crdenas-Pea, Andrs lvarez-Meza, Chester Dolph, Khan Iftekharuddin, Simon Eskildsen, Pierrick Coup, Vladimir Fonov, Katja Franke, Christian Gaser, Christian Ledig, Ricardo Guerrero, Tong Tong, Katherine Gray, Elaheh Moradi, Jussi Tohka, Alexandre Routier, Stanley Durrleman, Alessia Sarica, Giuseppe Fatta, Francesco Sensi, Andrea Chincarini, Garry Smith, Zhivko Stoyanov, Lauge Srensen, Mads Nielsen, Sabina Tangaro, Paolo Inglese, Christian Wachinger, Martin Reuter, John Swieten, Wiro Niessen, and Stefan Kleina. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge. In *NeuroImage*, volume. 111, pages 562- 579, 2015.

[15] Clive Holmes. Dementia. *Psychiatric Disorders*, volume. 40, pages 628- 631, 2012.

[16] David Munoz, and Howard Feldman. Causes of Alzheimer's disease. In *Canadian Medical Association or Its Licensors*, volume.162, pages 65-72, 2000.

[17] Esther Bron, Marion Smits, Wiro Niessen, and Stefan Klein. Feature Selection Based on SVM Significance Maps for Classification of Dementia. In *Machine Learning in Medical Imaging*, pages 272- 279, 2014.

[18] Fermn Segovia, Christine Bastin, Eric Salmon, Juan Manuel Grriz, Javier Ramrez, and Christophe Phillips. Combining Neuropsychological and neuroimaging data to assist the Early Diagnosis of Dementia. In *The 20TH Annual meeting of the Organization for the human brain Mapping*, 2014.

[19] German Berrios. Delirium and Confusion in the 19th century: a Conceptual History. In *British Journal of Psychiatry*, volume. 139, pages 439- 449, 1981.

[20] Guiomar Nascimento Oliveira, Carolina Pinto Souza, Maria Foss, and Vitor Tumas. An analysis of the cognitive items of the movement disorders society checklist for the diagnosis of dementia in patients with Parkinson's disease. In *Parkinsonism & Related*

Disorders, volume. 21, pages 1260- 1263, 2015.

[21] Ingrid Scholl, Til Aach, Thomas M. Deserno, and Torsten Kuhlen. Challenges of medical image processing.In *Springer*, volume 26, pages 5- 13, 2010.

[22] Heung-Il Suk and Dinggang Shen. Deep Learning-Based Feature Representation for AD/MCI Classification. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013 Lecture Notes in Computer Science*, pages 583- 590, 2013.

[23] J. Ramrez, J. Grriz, D. Salas-Gonzalez, A. Romero, M. Lpez, I. lvarez, and M. Gmez-Ro. Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features. In *Information Sciences*, volume 237, pages 59- 72, 2013.

[24] Kazunari Ishii. PET Approaches for Diagnosis of Dementia. In *American Journal of Neuroradiology*, volume 35, pages 2030- 2038, 2013.

[25] Lorenzo Rocchi, Flavia Niccolini, and Marios Politis. Recent imaging advances in neurology. In *Journal of Neurology*, volume 262, pages 2182- 2194, 2015.

[26] Lauge Sorenson, Cecilie Anker, Loana Balas, Martin Lillholm, Christian Igel, and Mads Nielsen. Dementia diagnosis using MRI cortical thickness, shape, texture, and volumetry. In *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pages 111- 118, 2014.

[27] Pierre Baldi. Autoencoders, Unsupervised Learning, and Deep Architectures. In *Journal of Machine Learning Research*, volume 27, pages 37- 50, 2012.

[28] Janine Stein, Melanie Luppa, Elmar Brahler, Hans-hemut Konig, Steffi Riedel-heller. Assessing cognitive changes in the elderly: Reliable Change Indices for the Mini-Mental State Examination. In *Acta Psychiatric Scandinavia*, volume 126, pages 208- 218, 2012.

[29] Eric M. Reiman, M.D.1 and William J. Jagust. Brain imaging in the study of Alzheimer's disease. In *NeuroImage*, volume 61, pages 505- 516, 2012.

[30] Karen Choong, Rajasvaran Logeswaran, and Michel Bister. Improving Diagnostic Quality of MR Images Through Controlled Lossy Compression Using SPIHT. In *Journal of Medical Systems*, volume 30, pages 139-143, 2006.

[31] Walter Kukull, Eric Larson, E., Reifler, T. Lampe, M. Yerby, and Julie Hughes. Interrater reliability of Alzheimer's disease diagnosis. In *Neurology*, volume 40(2), pages 257- 260, 1990.

[32] Andrea Mechelli, Cathy Price, Karl Friston, and John Ashburner. Voxel-Based Morphometry of the Human Brain: Methods and Applications.In *Current Medical Imaging Reviews CMIR*, volume 1(2), pages 105- 113, 2005.

[33] Alfonso Nieto-Castanon, Satrajit Ghosh, Jason Tourville, and Frank Guenther. Region of interest based analysis of functional imaging data. In *NeuroImage*, volume 19(4), pages 1303- 1316, 2003.

[34] Claire Boutet, Marie Chupin, Stphane Lehricy, Linda Marrakchi-Kacem, Stphane Epelbaum, Cyril Poupon, Christopher Wiggins, Alexandre Vignaud, Dominique

Hasboun, Bndicte Defontaines, Olivier Hanon, Bruno Dubois, Marie Sarazin, Lucie Hertz-Pannier, and Olivier Colliot. Detection of volume loss in hippocampal layers in Alzheimer's disease using 7 T MRI: A feasibility study. In *NeuroImage: Clinical*, volume 5, pages 341- 348, 2014.

[35] Vanessa Guzman, Owen Carmichael, Christopher Schwarz, Giuseppe Tosto, Molly Zimmerman, and Adam Brickman. White matter hyperintensities and amyloid are independently associated with entorhinal cortex volume among individuals with mild cognitive impairment. In *Alzheimer's & Dementia*, volume 9, 2013.

[36] Samuel Barrack. Biochemical biomarkers in Alzheimer's disease. In *Internet medical publishing*, 2013.

[37] C. Li, Spectral Geometric Methods for Deformable 3D shape Retrieval, M.S. thesis, Information System Engineering, Concordia University, Montreal, Quebec, Canada, 2013.

[38] Erik Musiek, Yufen Chen, Marc Korczykowski, Babak Saboury, Patricia Martinez, Janet Reddin, Abass Alavi, Daniel Kimberg, David Wolk, Per Julin, Andrew Newberg, Steven Arnold, and John Detre. Direct Comparison of FDG-PET and ASL-MRI in Alzheimer's disease. *Alzheimer's dement*, volume 8(1), pages 51- 59, 2012.

[39] Eliasz Engelhardt, Carla Tocquer, Charles Andr, Denise Madeira Moreira, Ivan Hideyo Okamoto, and Jos Luiz de S Cavalcanti. Vascular dementia: Diagnostic criteria and supplementary exams: Recommendations of the Scientific Department of Cognitive Neurology and Aging of the Brazilian Academy of Neurology, Part I. In *Dementia & Neuropsychological Dement.* Neuropsychology, volume 5 (4), pages 251- 263, 2011.

[40] Paul Kinahan and James Fletcher. Positron Emission Tomography Computed Tomography Standardized Uptake Values in Clinical Practice and Assessing Response to Therapy. In *Seminars in Ultrasound, CT and MRI*, volume 31(6), pages 496- 505, 2010.

[41] Bharti Rana, Akanksha Juneja, Mohit Saxena, Sunita Gudwani, S. Senthil Kumaran, Madhuri Behari, and R. K. Agrawal. Graph-theory-based spectral feature selection for computer aided diagnosis of Parkinson's disease using T1-weighted MRI. In *Inter-*

national Journal of Imaging System and Technology, volume 25(3), pages 245-255, 2015.

[42] C. B. Levine, K. R. Fahrbach, A. D. Siderowf, R. P. Estok, V. M. Ludensky, and S. D. Ross, Diagnosis and Treatment of Parkinson's disease: A Systematic Review of the Literature. In *Evidence Report/Technology Assessment*, Volume 57, pages 1- 4, 2003.

[43] Jenn Riley. Understanding metadata.In *NISO primer*, (2004).

[44] Li Deng and Dong Yu. Deep Learning: Methods and Applications. In *Foundations and Trends in Signal Processing*, volume 7, pages 197-387, 2013.

[45] Yichuan Tang. Deep Learning Using Linear Support Vector Machines. In *International Conference of Machine Learning*, Atanta, Georgia, USA, 2013.

[46] Daniele Ravi, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang Yang. Deep learning for Health informatics. In *IEEE Journal of Biomedical and Health Informatics*, volume 21, pages 4-21, 2017.

[47] Esther Bron, Marion Smits, Wiro Niessen, Stefan Klein. Feature Selection Based on the SVM Weight Vector for Classification of Dementia. In *IEEE Journal of Biomedical and Health Informatics*, volume 19, pages 1617-1626, 2015.

[48] Lesley Fellows, Howard Bergman, Christina Wolfson, and Howard Chertkow. Can clinical data predict progression to dementia in amnestic mild cognitive impairment. In *The Canadian Journal od Neurological Sciences*, volume 35(3), pages 314- 320, 2008.

[49] Garry Smith, Zhivko Stoyanov, Danica Greetham, Peter Grindrod, and James Saddy. Towards the Computer-aided Diagnosis of Dementia based on the geometric and network connectivity of structural MRI data.In *Proc MICCAI Workshop Challenge on Computer-Aided Diagnosis of Dementia Based on Structural MRI Data*, pages 101-110, 2014.

[50] All about Dementia. (2016). [Online]. Available: http://www.alzheimersresearchuk.org/wp-content/uploads/2015/01/All-about-dementia.pdf.

[51] Benefits of early dementia diagnosis. (July 2015). [Online]. Available: http://www.nhs.uk/conditions/dementia-guide/pages/dementia-early-diagnosis-benefits.aspx

[52] Allen Huang and Raymond Wu. (2016). Deep learning for music. [Online].Available:https://cs224d.stanford.edu/reports/allenh.pdf

[53] Douglas Eck and Jurgen Schmidhuber..*A First Look at Music Composition using LSTM Recurrent Neural Networks.*

[54] Brody Huval, Adam Coates, and Andrew Ng.Deep learning for class-generic object detection. In *Proceedings of the 31st International Conference on Machine Learning*,volume 32, 2013.

[55] Emdad Hossain and Girija Chetty. Multimodal Feature Learning for Gait Biometric Based Human Identity Recognition. In *International Conference on Neural Information Processing. Lecture Notes in Computer Science*, volume 8227, (2013).

[56] 10 Facts on dementia (2015) [Online]. Available: http://www.who.int/features/factfiles/dementia/en/.

[57] Martin Prince, Andres Wimo, Maelenn Guerchet, Gemma-Claire Ali,Yu-Tzu Wu, and Matthew Prina. (2015, Aug.). World Alzheimer report2015: The global impact of dementia. Alzheimer's disease international. London Available: ¡https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf.

[58] Michael Russell. (2012). Genetics of dementia [Online]. Available: $https : //www.alzheimers.org.uk/info/20010/risk_factors_and_prevention/117/genetics_of_dementia$

[59] Emma Reynish, L. Fratiglioni, Martin Prince, H. Bickel, A. Kiejna, and J. Georges. (2009). European Collaboration on Dementia - Prevention of dementia - Environmental/nutritional risk factors [Online]. Available: http://www.alzheimer-europe.org/Research/European-Collaboration-on-Dementia/Prevalence-of-dementia.

[60] What is Alzheimer's disease. (2012) [Online]. Available: $http : //www.usagainstalzheimers.org/crisis?gclid =$

$Cj0KEQiA6bq2BRC6\%20pp\%20f0_83Z1YIBEiQAgPYNvW\%20aCYhN4mgsGzzfTLIkXizAfH$

[61] Alzheimer Society of Canada. Other dementias. Internet: http://www.alzheimer.ca/en/About-dementia/Dementias

[62] Amyloid Plaques And Neurofibrillary Tangles (2016) [Online]. Available: http://www.brightfocus.org/alzheimers/infographic/amyloid-plaques-and-neurofibrillary-tangles.

[63] H. Knipe, and J. Jones. (n.d.). Vascular dementia [Online]. Available: http://radiopaedia.org/articles/vascular-dementia.

[64] *Information about FTD and dementia* (2016) [Online]. Available: https://www.ucl.ac.uk/drc/ftdsupport/information-about-ftd.

[65] *10 Early Signs and Symptoms of Alzheimer's* (2016) [Online]. Available: $http://www.alz.org/alzheimers_disease_10_signs_of_alzheimers.asp.$

[66] *Tests for Alzheimer's Disease and Dementia* (2016) [Online]. Available: $http://www.alz.org/alzheimers_{d}isease_{s}teps_{t}o_{d}iagnosis.asp - physical$

[67] Mayo clinic staff. (n.d.). *Dementia* [Online]. Available: http://www.mayoclinic.org/diseases-conditions/dementia/diagnosis-treatment/diagnosis/dxc-20198511

[68] *MRI scans* (n.d.) [Online]. Available: https://www.nlm.nih.gov/medlineplus/mriscans.html

[69] *Tests used in diagnosis of dementia* (n.d.) [Online]. Available: https://www.fightdementia.org.au/national/about-dementia/how-can-i-find-out-more/tests-used-in-diagnosing-dementia

[70] *Comparing types of imaging* (2015) [Online]. Available: http://www.nps.org.au/medical-tests/medical-imaging/for-individuals/imaging-compared.

[71] *Difference between MRI and PET scan* (2015 Sep.) [Online]. Available: http://theydiffer.com/difference-between-mri-and-pet-scan/

[72] *Health Care Professionals and Alzheimer's* (2016) [Online]. Available: http://www.alz.org/health-care-professionals/dementia-diagnosis-diagnostic-tests.asp.

[73] Wayne LaMorte. (2016 Jun.). Positive and negative Predictive Value [Online]. Available: $http : //sphweb.bumc.bu.edu/otlt/MPH - Modules/EP/EP713_Screening/EP713_Screening5.html$

[74] Rui DeFigueiredo, W. Rodman Shankle, Andrea Maccato, Malcolm Dick,Prashanth Mundkur, Ismael Mena, and Carl Cotman. Neural-network-based classification of cognitively normal, demented, Alzheimer disease and vascular dementia from single photon emission with computed tomography image data from brain. In *Proc. Natl. Acad. Sci.* U S A, volume. 92, pages. 5530-5534, 1995.

[75] Multiclass and multilabel algorithms (2012) [Online]. Available: $http : //scikit - learn.org/stable/modules/multiclass.html$.

[76] Binary Classification (n.d.) [Online]. Available: $http : //www.cse.iitk.ac.in/users/se367/10/presentation_local/BinaryClassification.html$.

[77] David Crdenas-Pea, Diego Collazos-Huertas, and German Castellanos-Dominguez. Centered Kernel Alignment Enhancing Neural Network Pretraining for MRI-Based Dementia Diagnosis. In *Computational and Mathematical Methods in Medicine*, 2016.

[78] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. In *Macmillan Publishers limited*, volume. 521, pages.436-444, 2015.

[79] Alina Kloss. Object detection using deep learning, 2015.

[80] Ashisha Gupta, Murat Ayhan, and Anthony Maida. Natural image bases to represent neuroimaging data. In *JMLR Workshop and Conference Proceedings*, vol. 28, pages.987-994, 2013.

[81] David Stutz. Understanding Convolutional neural networks, Seminar report, 2014.

[82] Analytics Vidhya, 2016. A comperhensive guide to data exploration. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/

[83] S. Sarraf et al., 2016. DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI. http://biorxiv.org/content/biorxiv/early/2016/08/21/070441.full.pdf.

[84] Mario Koppen. (n.d.). The curse of dimensionality. [Online]. Available: http://www.yaroslavvb.com/papers/koppen-curse.pdf

[85] Christos Davatzikos, Stephen Resnick, Xuesong Wu, P. Parmpi, and C. Clark. Individual patient diagnosis of AD and FTD via High- dimensional pattern classification of MRI. In *journal of Neuroimage*, vol. 41, pages. 1220-1227, 2008.

[86] Siqi Liu, Sidong Liu, Weidong Cai, Sonia Pujol, Ron Kikinis, and Dagan Feng. Early diagnosis of alzheimer's disease with deep learning. In *2014 IEEE 11th International Symposium on Biomedical imaging (ISBI)*, pages. 1015-1018. IEEE, 2014.

[87] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In *Neural Networks: Tricks of the Trade*, pages. 437-478, 2012.

[88] Adrian Rosebrock, 2016. http://www.pyimagesearch.com/2016/08/01/lenet-convolutional-neural-network-in-python/.

[89] World Health Organization. Fact Sheet on Dementia. 2016. [Online]. Available from: http://www.who.int/mediacentre/factsheets/fs362/en/

[90] Manhua Liu, Daoqiang Zhang, Dinggang Shen, and the Alzheimer?s Disease Neuroimaging Initiative. Hierarchical Fusion of Features and Classifier Decisions for Alzheimer's Disease Diagnosis. In *Human Brain Mapping*, vol. 35, pages. 1305-1319, 2013.

[91] Saima Farhan, Muhammad Abuzar Fahiem, and Huma Tauseef. An Ensemble-of-Classifiers Based Approach for Early Diagnosis of Alzheimer's Disease: classification using Structural Features of Brain Images. In *Computational and Mathematical Methods in Medicine*, 2014.

[92] Josef Kittler, Mohamad Hanef, Robert Duin, and Jiri Matas. On Combining Classifiers. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, No. 3, 1998.

[93] Dementia statistics. Available at http://www.alz.co.u mk/research/statistics

[94] Herve Abdi and Lynne Williams. Principal Component Analysis, 2010. [Online]. Available: http://www.utdallas.edu/herve/abdi-awPCA2010.pdf

[95] Jieping Ye. Least squares Linear Discriminant Analysis. In *ICML*, pages. 1087-1093, 2007.

[96] Nazri Mohd Nawi, Abdullah khan, M. Z. Rehman, Maslina Abdul Aziz, Tutut Herawan, and Jemal H. Abawajy. An Accelerated Particle Swarm Optimization Based Levenberg Marquardt Back Propagation Algorithm. In *International Conference of Neural Information Processing ICONIP*, pages. 245-253, 2014.

[97] What is OASIS. Internet: http://www.oasis-brains.org [May. 5, 2016].

[98] Satoshi Minoshima, Kirk A Frey, Robert A Koeppe, Foster, and Kuhl. A diagnostic approach in Alzheimer's disease using three-dimensional sterotactic surface projections of fluorine-18-FDG PET. In *J Nucl Med*, vol. 36, pages. 2030- 2038, 1995.

[99] http://www.fil.ion.ucl.ac.uk/spm/

[100] http://www.freesurfer.net/fswiki/MacOsInstall

[101] https://surfer.nmr.mgh.harvard.edu/fswiki/DownloadAndInstall

[102] Elham Alkabawi , Allaa Hilal, and Otman Basir. Feature Abstraction For Early Detection of Multi-type of Dementia with Sparse Auto-encoder. In *The 2017 IEEE International Symposium Conference on Systems, Man, and Cybernetics*, 2017.

[103] Robi Polikar, Apostolos Topalis, Devi Parikh, Deborah Green, Jennifer Frymiare, John Kounios, and Christopher M. Clark. An ensemble based data fusion approach for early diagnosis of Alzheimer's disease. In *Information Fusion*, vol. 9, pages. 83-95, 2008.

[104] Meiyan Huang et al. Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease. In *Scientific reports*, 2017.

[105] http://citeseerx.ist.psu.edu/viewdoc/download/doi=10.1.1.1.3162&rep=rep1&type=pdf

[106] Elham Alkabawi , Allaa Hilal, and Otman Basir. Computer aided classification of multi-type of dementia via Convolutional Neural Networks. In *the 12th Annual IEEE International Symposium on Medical Measurements and Applications*, July 2017.

[107] Jason Lerch. Measuring Cortical Thickness. Master thesis proposal, Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada, 2001. Available: http://www.bic.mni.mcgill.ca/users/jason/mastersproposal.pdf.

[108] Tanya Glozman and Orly Liba. Hidden Cues: Deep learning for Alzheimer's disease classification CS331B project final report.

Appendices

Appendix A

Appendix A: Publications

In This Thesis

Some of the research leading to this thesis has appeared previously in the following publications.

Journal Articles

- Alkabawi E, Hilal A, Basir O. **"A survey of Computer-Aided Vision Techniques for Early Diagnosis of Dementia"**. – *A Journal*, 2017.
- Alkabawi E, Hilal A, Basir O. **"A Context-aware CAD approach for early**

Identification of Multiple Types of Dementia ". – *A Journal*, 2017.

Conference Papers

- Alkabawi E, Hilal A, Basir O. **"Computer aided classification of multi-type of dementia via Convolutional Neural Networks"**. – *the 12th Annual IEEE International Symposium on Medical Measurements and Applications*, 2017, Minnesota, USA.

- Alkabawi E, Hilal A, Basir O. **"Feature abstraction for early detection of multi-type of dementia with Sparse Autoencoder"**. – *the 2017 IEEE International Symposium Conference on Systems, Man, and Cybernetics*, 2017, Banff, Canada.