COMPUTER AIDED DIAGNOSIS OF COGNITIVE DISORDERS USING
NEUROPSYCHOLOGICAL AND NEUROIMAGING DATA

by
Füsun Er

Submitted to Graduate School of Natural and Applied Sciences
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in
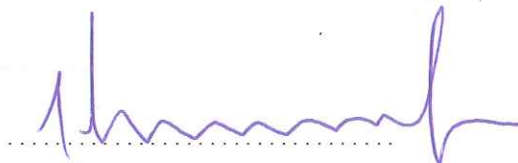Biotechnology

Yeditepe University
2017

COMPUTER AIDED DIAGNOSIS OF COGNITIVE DISORDERS USING
NEUROPSYCHOLOGICAL AND NEUROIMAGING DATA

APPROVED BY:

Assist. Prof. Dr. Dionysis Goularas
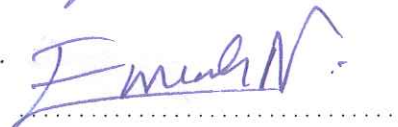(Thesis Supervisor)

Prof. Dr. Sibel Karşıdağ

Assoc. Prof. Dr. Emin Erkan Korkmaz

Assoc. Prof. Dr. Fatih Uğurdağ

Assist. Prof. Dr. Emrah Nikerel

DATE OF APPROVAL: ..../..../2017

ACKNOWLEDGEMENTS

ABSTRACT

COMPUTER AIDED DIAGNOSIS OF COGNITIVE DISORDERS USING NEUROPSYCHOLOGICAL AND NEUROIMAGING DATA

Computer-aided diagnosis (CAD) is an emerging area of neuroscience and neuroimaging that assists physicians in the interpretation of medical images. CAD is an interdisciplinary technology that combines artificial intelligence, computer vision, digital image processing and medical imaging. The overall goal of this thesis is to develop computer-aided diagnosis systems for patients with cognitive disorders based on neuropsychological and neuroimaging data using deep learning approaches. In this thesis, we developed three CAD systems. The first system aimed to distinguish diseases causing an abnormal decline in cognitive abilities that is not related to normal aging in an early stage, based on neuropsychological assessment data. The aim of the second CAD system was to predict the risk of developing Alzheimer's disease for mild cognitive impairment patients based on cross-sectional neuroimaging data. Finally, the third CAD system shared the same goal with the second CAD system, with a difference that it utilized longitudinal neuroimaging data. Furthermore, from a methodological perspective, we designed and implemented a novel pooling method for convolutional neural network models that is employed in the second and third CAD systems. Additionally, we designed and implemented a novel image preprocessing method for the longitudinal data of the third CAD system. The very high accuracy results achieved in this thesis has shown that deep learning technologies can be beneficial for diagnosis and predicting prognosis in cognitive disorders. Furthermore, CAD systems may contribute to better treatment of patients with cognitive disorders by altering their medication according to predicted prognosis.

ÖZET

NÖROPSİKOLOJİK VE NÖROGÖRÜNTÜLEME VERİLERİ İLE BİLİŞSEL BOZUKLUKLARIN BİLGİSAYAR DESTEKLİ TANISI

Bilgisayar destekli tanı (BDT), sinirbilimi ve beyin görüntüleme alanlarında ön plana çıkan, tıbbi görüntülerin yorumlanmasında doktorlara yardımcı olan uzman sistemlerdir. Ayrıca, yapay zeka, bilgisayar görüsü, dijital görüntü işleme ve medikal görüntüleme teknolojilerini bir araya getiren bir disiplinler arası çalışma alanıdır. Bu tezin genel amacı, derin öğrenme yaklaşımlarını kullanarak tıbbi görüntüleme verilerine dayalı bilişsel bozukluklar için bilgisayar destekli tanı sistemleri geliştirmektir. Bu amaç doğrultusunda, üç bilgisayar destekli tanı sistemi geliştirdik. Birinci çalışmada, nöropsikolojik değerlendirme verilerine dayanılarak, normal yaşlanmayla ilişkili olmayan bilişsel yeteneklerdeki anormal düşüşü erken aşamada ayırt etmek amaçlanmıştır. İkinci ve üçüncü bilgisayar destekli tanı sistemlerinin amacı, hafif bilişsel bozukluk hastalarında Alzheimer hastalığının gelişme riskini öngörmektir. İkinci bilgisayar destekli tanı sistemi kesitsel nörogörüntüleme verilerine dayanırken; üçüncü sistem, boylamsal nörogörüntüleme verilerinden faydalanmıştır. Ayrıca, bu tezde, konvolüsyonel sinir ağı modelleri için yenilikçi bir havuzlama yöntemi ve boylamsal veriler için yenilikçi bir görüntü ön işleme yöntemi önerdik. Çok yüksek doğruluğa sahip sonuçlar göstermiştir ki, derin öğrenme teknolojileri, hafif bilişsel bozukluğun prognozunu öngörmede başarılıdır. Ayrıca bu çalışma, bilgisayar destekli tanı sistemleri ile prognozu öngörülen hastalığa uygun ilaç tedavilerinin başlanması ile daha iyi bir tedaviye katkıda bulunulabileceğini göstermiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

ACC	Accuracy
AD	Alzheimer's disease
ADNI	Alzheimer's disease neuroimaging initiative
ANOVA	One-way analysis of variance
ARCD	Age-related cognitive decline
BNT	Boston naming test
CAD	Computer-aided diagnosis
CDR	Clinical dementia rating scale
C-MCI	Converted-mild cognitive impairment
CNN	Convolutional neural network
CSF	Cerebrospinal fluid
CT	Computed tomography
DBM	Deep Boltzmann machines
F1	F-measure
FDG	Fluorodeoxyglucose
FWHM	Full-width half-maximum
GLM	General linear model
HVI	Hippocampal volumetric integrity
LSTM	Long short-term memory
M5P	M5-model trees
MCI	Mild cognitive impairment
MMSE	Mini mental state examination
MRI	Magnetic resonance imaging
MP-RAGE	Magnetization-prepared rapid acquisition gradient echo
MTA	Medial temporal lobe atrophy
NC-MCI	Non converted-mild cognitive impairment
NCTB	Neurocognitive test battery
PET	Positron emission tomography
PPV	Precision
SD	Standard deviation

SMC	Significant memory concern
SPC	Specificity
SPM	Statistical parametric mapping
SVM	Support vector machine
TBM	Tensor-based morphometry
TPR	Sensitivity
VBM	Voxel-based morphometry
VD	Vascular dementia
VOI	Volume of interest
WMS-R	Wechsler memory scale-revised
WM	White matter

1. INTRODUCTION

The aim of this thesis is to investigate the use of machine learning and deep learning technologies in the field of medicine. More particularly, we attempted to investigate the use of these technologies for the prognosis prediction purposes for cognitive disorders. In other words, we examined the possibility of predicting the prognosis of cognitive disorders based on artificial intelligence methods. The ultimate goal is to provide computer-aided diagnosis (CAD) systems for physicians that will help them decide about the prognosis of a disease in order to propose a proper medication to patients. At first, we attempted to see if some cognitive diseases can be separable with a high success rate based on neurocognitive test scores. Furthermore, we examined the possibility of predicting the prognosis for mild cognitive impairment (MCI) patients that either they will develop Alzheimer's disease (AD) or remain stable. In order to achieve our goals we used radiological data as an input to computer aided diagnosis systems, which is presented in this thesis.

In order to understand computer-aided diagnosis properly, it is good to mention about importance of radiology in the diagnosis and treatment of diseases. Radiology is a medical discipline that uses medical imaging technologies to diagnose and treat diseases. Magnetic resonance imaging (MRI) is a medical imaging modality that uses strong magnetic fields and radio waves to acquire three-dimensional images of inside of the human body. The magnetic resonance (MR) scanning is a beneficial technology to physicians for diagnosing and determining treatment options in brain diseases (e.g. multiple sclerosis [1], Parkinson's disease [2], Alzheimer's disease [3]), cancer (e.g. prostate cancer [4], breast cancer [5], brain tumours [6]), diseases of other organs (e.g. liver [7], pancreas [8] or heart diseases [9]). A computed tomography (CT) scanner is a kind of X-ray machine that takes many X-ray images of different angles of the desired section of the body and produces tomographic images using computer programs [10]. The computed tomography is a powerful technology for imaging bone structures. Since a computed tomography scanner produces images faster than a magnetic resonance imaging scanner, it is widely used imaging technology in emergency services to detect bone injuries, brain traumas, chest problems. The ability to get benefit of medical imaging technologies depends on the acquisition quality of the images, as well as, on the interpretation quality of the images. A radiologist tries to catch visual clues

to interpret medical images for diagnosis or prognosis. Several experiments show that the radiologist misses the 20%-30% of those visual clues or may have difficulty to interpret the gathered visual clues [11]. Thus, the computer-aided diagnosis is an interdisciplinary study to develop computer systems to assist the radiologist in catching or interpreting visual clues of medical imaging [12], which combines several technologies like artificial intelligence, computer vision, digital image processing to process several medical imaging modalities for specific diagnostic or prognostic purposes. In this thesis, we developed three CAD systems in order to help physicians in predicting the diagnosis or prognosis of cognitive disorders.

1.1. MOTIVATION

The motivation behind our thesis comes from two perspectives. The first one is the biological perspective, which is the motivation behind the selected disease to be investigated in this thesis. In other words, this perspective defined the diagnostic or prognostic goals of the computer-aided diagnosis models. The second one states the motivation about which methods to be used in the development of the computer-aided diagnosis models to achieve our goals.

1.1.1. Biological Perspective on Motivation

In this thesis, we decided to predict diseases that are related with the cognitive impairments of elderly people. The reason lies on the fact that this type of diseases, more specifically dementias, are widespread and a successful prognosis can be beneficial for individuals but also for state institutions in terms of finance.

Recent community studies indicated that the prevalence of dementias is rising all over the world in the last decade [13]. The ability to take care of themselves may decrease day-by-day in the patients with dementias, and, they need a growing healthcare. This means that, dementia has a great economical, physical and psychological burden on not only individual, but also for their family and society [14]. According to the World Alzheimer's report of 2015, over 46 million people currently living with dementias in the world, and, it is estimated that the number of people suffering from dementias will increase up to 131.5 million by the year of 2050. Nowadays, the total estimated worldwide cost of dementias is estimated as

$818 billion [15]. The psychological and physical impact of dementias on the caregivers and families of patients with dementias can not be ignored. Many family carers, who are caring a family member with dementia, suffer from several psychological illness like depression, anxiety and feelings of exhaustion due to providing an intensive support to the patient including making difficult treatment decisions, dealing with the financial problems, providing personal care and housekeeping of the patient [16].

Beside the psychological impacts of dementias, taking care of a family member with dementia may also cause several physical conditions like impaired immunity, high blood pressure and cardio-vascular diseases, which are generally caused by decreased mobility due to isolation from social life, interrupted sleep due to night-time care and exhausting work load [16].

Despite its prevalence and importance, the term of dementia is generally confused with the term of Alzheimer's disease. The dementia refers to a group of symptoms that are characterized by a decline in cognitive abilities. Those cognitive deficits associated with dementias are severe enough to disrupt a normal daily life. Among dementias, the Alzheimer's disease accounts for up to 75% of all dementia cases, which is a progressive neurodegenerative disorder [17]. The dementia caused by conditions related with an inadequate blood flow to the brain is called as vascular dementia (VD), which is considered as the second most common type of dementias with a prevalence percentage of 30% [18]. However, many other conditions like brain tumours [19], diabetes [20], thyroid disorders [21], depression [22] and long-term use of alcohol [23] may cause similar symptoms with dementias.

The neurodegeneration is initially characterized by synaptic damage and followed by neuronal loss in Alzheimer's disease [24]. In the course of time, as a result of neurodegeneration, brain tissue shrinks in volume, thus causing cognitive impairment. On the other hand, mild cognitive impairment is a clinical diagnosis indicating abnormal decline in cognitive abilities more than expected for normal aging. And, it has been associated with a significant risk of developing Alzheimer's disease. Although mild cognitive impairment is characterized by a decline in memory. In many cases, the memory problems experienced in mild cognitive impairment may remain stable for years, unlike Alzheimer's disease.

1.1.2. Methodological Perspective on Motivation

In this thesis, we decided to use machine learning methods for developing prediction systems for dementias. The motivation lies on the fact that medical data related with dementias are huge in size and present an important complexity and variety; luckily, artificial intelligence methods are providing the necessary platform for solving such problems related with the prediction of a disease.

In the machine learning studies, the first and most important step is defining features of the system. In that context, a feature is defined as a measurable property of an object or a phenomenon being observed. Actually, the success of a machine learning algorithm is strongly depended on the feature definition [25]. This fact has raised a study field called as *feature engineering* which deals with the representation of the data as a set of features that is computationally convenient to process using several methods for extracting, transforming or selecting them [26]. A well designed feature set is essential for a machine learning model in order to reduce the complexity of the model, to increase the accuracy and to prevent overfitting.

Traditional feature engineering in neuroimaging area is a manual process that requires the domain knowledge of the data, like defining region-of-interests of white matter hyperintense lesions [27], amyloid deposition [28] or other amyloid imaging biomarkers [29]. Even if an automated tool is used, this kind of hand-crafted features may need to be verified by an domain expert; like determining subtype of white matter hyperintense lesions [30], region-of-interest of ventricles [31].

The traditional feature extraction procedure in neuroimaging studies is handled in several ways. Considering all voxel intensities of the neuroimaging data as a feature set is a widely used technique for the classification purposes [32]–[35]. Since the neuroimaging data is high dimensional, several studies employed feature selection methods to reduce the dimension of the feature vectors. A special scalar index was proposed for selecting a set of voxels to train a classifier algorithm [36]. In another study, a subset of voxels was selected based on a proposed scoring algorithm called as structural abnormality index score [34]. Moreover, the volume of specific brain tissue [37]–[39] by means of tissue segmentation algorithms or

brain regions [40]–[43] were taken into account. Several studies measured cortical thickness of specific parts of the brain to be used as a feature vector for constructing a diagnostic model of Alzheimer's disease [44], [45]. Gerardin *et al.* proposed a method for extracting shape features of the hippocampus to be used in CAD systems [46].

The issue of extracting convenient and efficient features of neuroimaging led us to search for automatic feature extraction methods. In this context, deep learning has attracted our attention, which is an emerging research field of machine learning for learning automatic representation of data. Hence, these facts have shaped our motivation for this thesis to explore applicability of deep learning techniques for the automation of feature extraction of neuroimaging.

1.2. OBJECTIVES

The first aim of this thesis is to build computer-aided diagnosis systems for cognitive disorders and to assess the applicability of deep learning techniques to learn features from the high-dimensional neuroimaging data. For this purpose, both baseline and 12-months follow-up structural MR images of patients with a baseline diagnosis of mild cognitive impairment were acquired from the Alzheimer's disease neuroimaging initiative (ADNI) dataset, which it is known that they are either remained stable in mild cognitive impairment state or eventually developed Alzheimer's disease during their follow-up. In this thesis, convolutional neural networks (CNN) were employed for the automatic feature extraction procedure. The prognosis predictive power of the acquired features was assessed using traditional machine learning algorithms.

The second aim of this thesis is to build computer-aided diagnosis systems for age-related cognitive decline based on neuropsychological test scores. In order to realize this goal, we acquired the neuropsychological assessment data of patients with age-related cognitive decline, Alzheimer's disease, vascular dementia and mild cognitive impairment from the neurology department of Maltepe University Hospital. In this thesis, we built a computer-aided diagnosis system for dementia, which is utilizing statistical methods and machine learning algorithms based on the neuropsychological test scores of the patients.

1.3. OUTLINE OF THE THESIS

This thesis is divided into five chapters including this current chapter. The contents of each chapter are presented in detail below.

Chapter 1 is the current chapter where the introduction and motivation of this thesis is presented.

Chapter 2 provides necessary background information of the important concepts used in this thesis.

In Chapter 3, a detailed explanation of the applied research methodology is presented.

Chapter 4 presented the findings and results of the conducted experiments.

Finally, we concluded and discussed the results and stated future works in Chapter 5.

2. BACKGROUND

This chapter provides some general conceptual background about machine learning, deep learning, voxel-based morphometry and neuroimaging modalities for a better understanding of the methodology of this thesis that are explained in the Chapter 3. Furthermore, the Alzheimer's Disease Neuroimaging Initiative is introduced, which is the institution we obtained MRI data for this thesis. Finally, recent computational approaches are given in terms of diagnostic goals, feature representation methodologies and the use of convolutional neural networks for neuroimaging.

2.1. GENERAL INFORMATION

The concept of machine learning, deep learning, voxel-based morphometric analysis and neuroimaging modalities are given from historical perspective in this section.

2.1.1. Machine Learning

The discipline of machine learning deals with the question of how a computer programs themselves for a specific task and gain experience without implicitly programmed. What we mention for the term of machine learning is that the learning algorithms for a computer to find one suitable solution among all possible solutions itself by learning from the data.

From the historical perspective, the checker program developed by Arthur Samuel is considered as the first machine learning application in the 1950s [47]. Before this, McCulloch and Pitts developed the first mathematical model of an artificial neuron known as McCulloch-Pitts neuron that is inspired from the human brain [48]. The model was working as a threshold logical unit, where simple logical functions, like OR and AND, can be implemented with a single McCulloch-Pitts neuron. In 1949, Donald Hebb introduced the concept of synaptic plasticity that the synaptic strength is not constant in the brain, more specifically, simultaneously activating cells strengthen the synapses [49]. This discovery in neuroscience pioneered the development of a new artificial neuron model, called as perceptron, which is invented by Frank Rosenblatt in 1957 [50]. Marvin Minsky and

Seymore Papert revealed the weakness of perceptrons, stating that the perceptrons solve only linearly separable problems. With the development of multi-layered perceptron networks and the use of back-propagation algorithm for dealing with the problem of training hidden layers, the field of artificial neural networks gained popularity again. In 1995, a new learning algorithm, named as support vector machine (SVM), is proposed by Vapnik *et al.* [51]. Nowadays, the SVM is still widely studied learning algorithm in the field of machine learning.

2.1.2. Deep Learning

The concept of deep learning is a new area of machine learning concerning hierarchical representations of the data, where the higher level features are derived from the lower level features. As many researches agree on that, the history of deep learning starts with the study of McCulloch and Pitts in 1943 [52]. The studies of Ivakhnenko and Lapa were considered as the earlier applications of the deep structures in 1965 [53]. The model introduced by Kunihiko Fukushima in the 1980s, called as Neocognitron, inspired the convolutional neural networks [54]. Thus, LeCun *et al.* proposed the convolutional neural network model with the implementation for the recognition of handwritten zip code. With the development of high-speed computers and huge amount of memories, the studies of deep learning methods gained speed in the 2000s. The long short-term memory (LSTM) was proposed by Hochreiter and Schmidhuber in 1997 [55], Hinton introduced deep belief networks (DBN) [56], and Salakhutdinov and Hinton proposed deep Boltzmann machines (DBM) [57] in 2009. Recently, the use of deep learning methods have gained popularity in the neuroscience studies [52].

In recent studies, two common buildings blocks were used to construct deep networks structures: autoencoder and restricted Boltzmann machine (RBM). The main purpose of these building blocks are to learn the latent representation of the data. Although both share the similar idea, they have differences in their structures. Such that, an autoencoder is a kind of feed-forward artificial neural network [58] and a restricted Boltzmann machine is a stochastic neural network [59]. A restricted Boltzmann machine is composed of a set of bi-directionally connected binary visible units in one visible layer and a set of binary hidden units in one hidden layer [59]. Unlike an autoencoder, in a RBM each visible unit is bi-

directionally connected to all hidden units. Also they differ in the training procedure, the back propagation algorithm is utilized to train an autoencoder, such that, the output data is almost same as the input data. Thus, the output of the hidden neurons in the trained autoencoder is the latent representation of data. On the other hand, the contrastive divergence [60] algorithm is used for training a restricted Boltzmann machine [61].

Deep networks are constructed using these building blocks. The most commonly used deep network is stacked autoencoders, which is composed of multiple layers of autoencoders [62]. Such that, the output of each outer layer is transmitted to the inner layer as input. Figure 2.1 shows an illustration of a stacked autoencoder with two hidden layers. The acquired latent representation of the data at each level is marked with red.
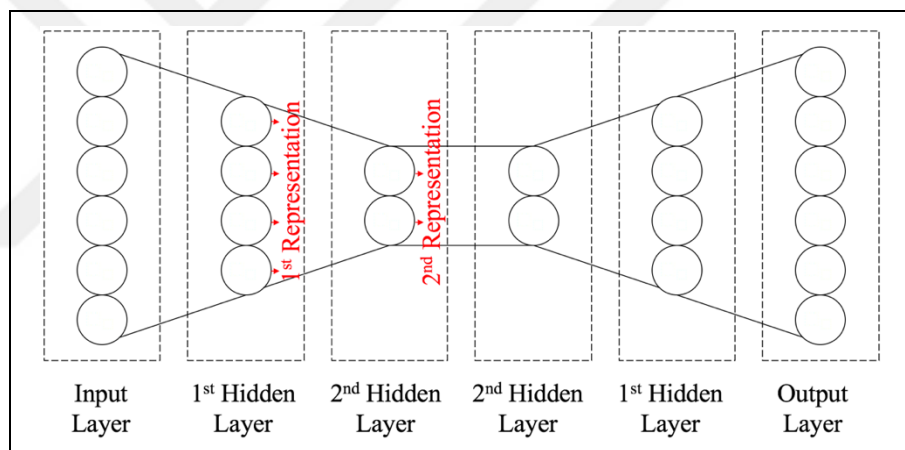
Figure 2.1. General schema of a two-layered stacked autoencoder

The training of a stacked autoencoder is handled as a stack of several single-layer autoencoders. Firstly, a simple autoencoder network is constructed for the outermost layer of the deep structure, which the number of hidden units is same as the number of hidden units in the first hidden layer of the deep network. Then, the autoencoder is trained to acquire the first latent representation of the data. The acquired latent representation becomes the input to the next inner layer. In the following step, a new autoencoder network is built, where the number of the input units is same as the size of the previously acquired latent representation and the number of hidden units is same as the number of hidden units in the second hidden layer of the deep network. Similarly, this second network provides the second latent representation of the data. In this way, the acquired latent representations are

propagated through the remaining hidden layers of the deep network. The same procedure is repeated as it is done for the outermost layer. The outmost layer produces the high-level representation of the input data.

Just as the stacked autoencoders are made up of a stack of autoencoders, the deep belief networks (DBN) are composed of a stack of restricted Boltzmann machines. Another similarity exists in the training of the network. In other words, the deep belief networks are trained layer-wise, too.

A convolutional neural network (CNN) is another kind of feed-forward neural network that is used to learn latent representation of grid-structured data, like images or time-series data. The concept of convolutional neural network was originally proposed by LeCun *et al.* and applied to recognize handwritten zip code. Nowadays, it is widely used technique in neuroimaging studies due to the fact that neuroimaging data are generally acquired and analysed in a series of images. A convolutional neural network is composed of a consecutive set of convolutional and pooling layer pairs.
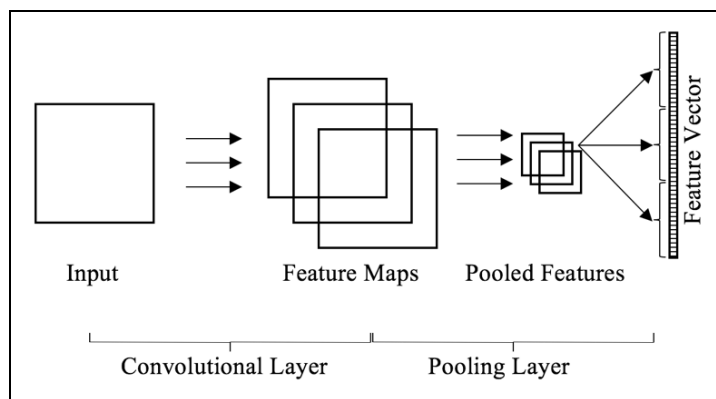
Figure 2.2. General schema of a single-layer convolutional neural network

To illustrate the concept of CNN, Figure 2.2 shows a general schema of a single-layer convolutional neural network, which composed of one pair of convolutional and pooling layer. In the convolutional layer, each input is convolved with a set of filter to construct a set of feature maps. Each feature map highlights different features of the input data depends on the corresponding filter. In the pooling layer, each feature map is down sampled individually, called as pooled features. Pooling layer is used in convolutional neural

networks to progressively reduce the spatial size of the representation in each layer. The main advantage of the pooling for the CNN model is the prevention of overfitting by reducing the size of feature maps. Finally, the acquired pooled features are vectorised to construct a feature vector, which represents the input data. In other words, this feature vector is the numeric representation of the image data that can be used further classification purposes to train a machine learning algorithm.

2.1.3. Voxel-based Morphometry

It is better to explain "what voxel-based morphometry (VBM) is" for a better understanding of this thesis. The VBM is a statistical method used in neuroimaging analysis to investigate brain tissue abnormalities [63]. The term of statistical parametric mapping (SPM, http://www.fil.ion.ucl.ac.uk/spm/) refers to the name of the technique developed by Friston *et al.* and the name of the software tool developed by the Wellcome Department of Imaging Neuroscience. In the terminology of SPM, the concept of VBM incorporates two main stages: data preprocessing stage and statistical analysis stage.

The data preprocessing stage of VBM is composed of *spatial normalization, segmentation, modulation* and *smoothing* steps. The goal of the *spatial normalization* is to align anatomical locations among subjects, which is achieved by registering images into the same template [64]. In the *segmentation stage*, the probabilities of being grey matter, white matter or cerebrospinal fluid for each voxel are calculated using a pre-defined tissue probability map (TPM), which TPM is provided by the International Consortium for Brain Mapping [65]. The spatial normalization and segmentation steps are unified in later versions of SPM software [66]. The aim of the *modulation* is to preserve local probabilistic volumes in spatially normalised tissue segmentation images, which is based on the change of variable theorem. Finally, it is recommended to smooth images to increase the sensitivity to detect changes, which each voxel is the weighted average of surrounding voxels in smoothed images [64].

In the statistical analysis stage, this preprocessed images are investigated group-wise statistically. In order to more specifically, a t-value or F-value is derived from statistical comparison between two groups of preprocessed images for each-voxels separately.

2.1.4. Neuroimaging Modalities

One of the most commonly used technologies by physicians to diagnose or make decision about the treatment of brain diseases is neuroimaging. The neuroimaging is a general term refers in vivo imaging techniques, which can be grouped into two main categories: structrural neuroimaging and functional neuroimaging. In this thesis, we dealt with structrural neuroimaging due to the relation between abnormalities in grey matter tissue and dementias.

Structural imaging technologies are aimed to get images of the anatomical structures of the inside of the body. The structural images are good at showing the contrast between different tissue types of the brain (e.g. grey matter, white matter and cerebrospinal fluid). Several technologies are used to acquire structural brain scans: computed tomography (CT) and magnetic resonance imaging (MRI).

The first technology to create structural brain images is the CT scanning of the brain. A CT scanner exposes a low level radiation to the brain in several directions. Then, the applied radiation passes through the brain tisue. The amount of radiation being absorbed differs in each tissue type due to the difference in densities. This is idea behind that how a CT scanner produces structural imaging of the brain.

Another kind of technology to produce structural images of the brain is magnetic resonance imaging. An MRI scanner uses a huge magnet and radio pulses to create images. Due to the existance of gigant magnetism, all protons in the brain are aligned. When a radio pulse is applied, some of the protons starts to spin in a specific frequence. After a while, the radio pulse is stopped and then the protons starts turn back to their initial states. During this relaxation period, each protons releases different amount of energy to be processed by computers for producing structural images of the brain. This technology is preferable to the CT technology due to lack of radiation and capability of producing high contrast images. On the other hand, a CT scanner produces structural images faster than an MRI scanner. This makes the CT technology favourite in emergency services, especially to detect the injuries and traumas. In this thesis, we prefered to use the structural neuroimages produced by the MRI scanners due to their high spatial resolution and contrast.

2.1.5. The Alzheimer's Disease Neuroimaging Initiative

We acquired our MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) for training our CAD models in this thesis. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is an ongoing, multi-center, multi-million dollar study launched in 2004 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Brain Imaging and Behaviour Administration (FDA), private pharmaceutical companies, and non-profit organizations. The scientists of ADNI at research centres in the U.S. and Canada have been collecting, validating and processing medical data of elderly cognitively normal controls and patients diagnosed with any stages of Alzheimer's disease. Demographic information, biospecimens and physical assessment data, MRI and PET scan data, APOE and whole-genome genotyping data, and neuropsychological test scores of the patients were collected at the baseline and follow-up stages in a predefined protocol. Those data are shared in world-wide with the neuroscientists through a data portal. Hereby, it is intended to support the research studies conducted to understand the progression of the pathology of Alzheimer's disease. The preliminary goal of the ADNI was to identify clinical, imaging, genetic, and biochemical biomarkers as predictors for Alzheimer's disease for use in clinical trials and diagnosis. Furthermore, with the help of the cutting-edge developments in neuroimaging, the ADNI studies encapsulated to understand the progress of cognitive decline in AD patients, to develop new methods as a result testing the effectiveness of alternative detective methods.

The follow-up data from participants of the longitudinal ADNI study were collected in three phases, named as: ADNI-1, ADNIGO, and ADNI-2. Some participants were carried forward to next phases for continued monitoring, as well as, new participants were added to further investigation for early and differential diagnosis of Alzheimer's disease, and for the tracking of AD progression.

The initial phase (ADNI-1) was launched in 2004 and continued till 2010. 819 participants , aged 55 to 90, were included in this phase, which consist of 229 cognitively normal elderly, 398 patients diagnosed by MCI in baseline, and 192 patients with early AD. The second stage of the ADNI study (AGNIGO) started in June 2009 with 500 Normal Controls and MCI participants of ADNI-1. 210 participants were moved to the ADNI-1 and 129 new

participants in the early mild cognitive impairment (EMCI) were enrolled during ADNIGO phase till 2011. With the newly enrolled EMCI participant, the researchers have opportunity to understand the characteristic of the AD that precedes MCI. MRI were collected in 3T on all newly enrolled subjects, while MRI follow-up scans were continue in 1.5T for transferred ADNI-1 participants [67]. Finally, the ADNI-2 is the third and on-going phase which launched in 2011. Some participants of ADNIGO and ADNI-1 were forwarded to ADNI-2. 186 Normal Control, 179 early MCI (EMCI), 163 late MCI (LMCI), and 148 mild AD have been newly enrolled in the ADNI-2 stage till now. A new cohort of the significant memory concern (SMC) patients were added in the ADNI-2 stage with 106 participants.
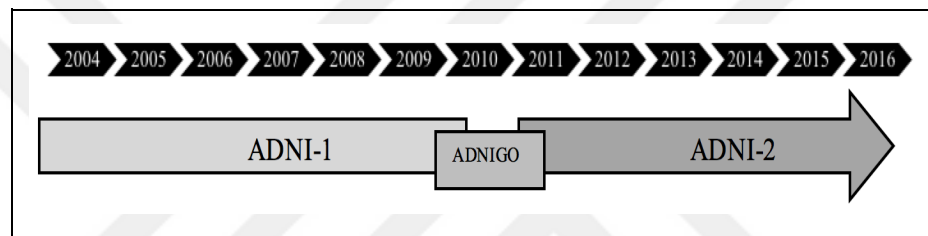
Figure 2.3. Timeline of three phases of the ADNI study

Figure 2.3 illustrates the timeline of the phases starting from the beginning of the ADNI-1. The ADNI-1 and ADNIGO phases were closed, but, the ADNI-2 phase still continues to enrol new participants and includes the follow-up of patients who enrolled in the previous phases.

Table 2.1. The number of the participants for each phases grouped by baseline diagnosis

	ADNI-1		ADNI-GO		ADNI-2		TOTAL
	transfer	new	transfer	new	transfer	new	
NC	-	229	95	-	126	186	415
AD	-	192	1	-	1	148	340
EMCI	-	-	-	129	120	179	308
LMCI	-	398	114	-	148	163	561
SMC	-	-	-	-	-	106	106
TOTAL	0	819	210	129	395	782	1730

Table 2.1 shows the number of the participants for each phase grouped by their baseline diagnosis (cohorts). The number of new enrolments occurred during the phase and the

number of transferred participants are given separately. Additionally, the total number of the phases and cohorts are provided separately. The grand total number of participants was 1730 when this analysis is executed.

Among all 1730 participants, 140 had a baseline diagnosis of MCI, had 3T T1-weighted MRI data through all their follow-up and they eventually developed AD. The number of patients in each phase and each convert time are presented in Table 2.2 for those 140 patients. In this thesis, we called them as C-MCI (converted-MCI) patients. Among all 140 patients, a total of 26 patients converted to AD six months later baseline. Similarly, of 38 patients after 12 months, of 11 patients after 18 months, of 39 patients after 24 months, of 18 patients after 36 months, of 5 patients after 48 months, of one after 72 months and of two after 96 months from the baseline examination.

Table 2.2. The number of patients grouped in phase and convert time

	CONVERT TIME AFTER BASELINE								TOTAL
	m06	m12	m18	m24	m36	m48	m72	m96	
ADNI-1	3	17	11	5	5	3	1	2	47
ADNI-GO	-	3	-	4	3	2	-	-	12
ADNI-2	23	18	-	30	10	-	-	-	81
TOTAL	26	38	11	39	18	5	1	2	140

Again among all 1730, 165 patients were had a baseline diagnosis of mild cognitive impairment and remained stable in mild cognitive impairment during their follow-up at least four follow-up periods. We called them as NC-MCI (non-converted-Mild Cognitive Impairment) patients through this thesis.

2.2. COMPUTATIONAL APPROACHES IN NEUROIMAGING

Machine learning is widely used techniques in neuroscience for classification of neurological diseases or conditions; like, Alzheimer's disease vs. healthy control (HC) [32]–[34], [68], AD vs. MCI [40], [69]–[73], AD vs. frontotemporal lobar degeneration [33] and schizophrenia (SC) vs. HC [74]. Among those kind of studies, several kind of data were utilized to train machine learning algorithms; like, neurosychological assessment data [75]–

[78], functional MRI [79], [80], structural MRI [70], [81], [82] or any combinations of them [83]–[87].

In processing neuroimaging by machine learning algorithms, a feature extraction procedure is needed to create a numeric representation of the neuroimaging data. The mostly used and easy method is to vectorize the data that each voxel will be one feature [32]–[35]. However, the problem of high-dimensionality is arised in that kind of representation. Several methods were employed to tackle with the high-dimensionality problem in the literature; like, considering only a set of voxels based on a specific kind of tissue type [37]–[39], based on a specific score defined for each voxel [34], [36].

Another way to prepresent neuroimaging data as a feature vector is to calculate grey matter volumes of anatomical parts of the brain [40]–[43]. Sparse auto-encoder [71] and deep Boltzmann machine [70], high-order graph matching [88], multi-modality canonical [89], tree-structured sparse learning method [90] and kernel-based methods [73] are another methods for reducing the high-dimensionality of the neuroimaging data.

One approach to learn kernels for the convolution layer of the convolutional neural networks is to train an autoencoder on randomly extracted patches of the dataset, that is proposed by Ng *et al.* [91]. Firstly, the patches are vectorised separately to generate a dataset for training the autoencoder. After training, the weight vectors of each hidden neuron of the trained autoencoder are reshaped in the same size as the patches. Therefore, this procedure results in as many images as the number of hidden neurons, which are to be used as kernels in the convolutional layer for generating the feature maps. These kernels are expected to detect the most common features of the data. This approach have been successfully applied to neuroimaging data by Gupta *et al.* [81], [82] for differentiating AD from MCI patients. Payan *et al.* [92] were followed the study of Gupta *et al.* [81], [82] for investigating the use of three-dimensional patches for the convolutional neural network [80].

The other approach for learning filters for the convolutional neural networks is to initialize kernels randomly in the beginning and train the convolutional neural network with backpropagation algorithm. For this purpose, the convolutional neural network structure should include supervised classifier at the last layer [93]. In other words, this approach is a

supervised learning procedure that should need the class membership information of the data.

There are several kind of pooling methods used for the pooling layer of convolutional neural network in the literature. The widely used types of pooling include the max-pooling, min-pooling and average-pooling methods. Those three methods are similar during the region selection for pooling. In this method, a window is slided over the image for resampling purposes, in combination with a scalar value that defines how much the window should be shifted each time. Then, according to the applied method, the maximum, minimum or average value on the window is taken as the output at every shift. Other pooling methods proposed in the literature are rank-based [94], adaptive spatial [95], attention pooling-based [96], geometric l_p-norm [97], fractional-max [98], stochastic [99], combined local and global pooling [100] and multiple spatial pooling [101] approaches. In rank-based pooling, the rank of activation in a pooling region is utilized to increase the discriminant ability of the pooled features [94]. Liu *et al.* calculated a spatial distribution matrix to define how the image patches are pooled together [95]. The new pooling scheme, called as attention pooling, combines the convolutional layer and the bidirectional long-short term memory [96]. The geometric l_p-norm pooling method is preserving the class-specific spatial/geometric information in the pooled features [97]. Graham *et al.* proposed a fractional version of max-pooling to construct suitable pooling regions [98], which is a kind of stochastic pooling [99]. Both local and global pooling strategies were combined in the study of Xiong *et al.* [100]. Finally, the features were pooled according to the relations between features and Gaussian distributions [101]. To sum up, the pooling algorithms are the algorithms that defines which data points will be pooled together to form pooled features.

2.3. SUMMARY OF THE CHAPTER

In this chapter, we provided the necessary background information on machine learning, deep learning, voxel-based morphometric analysis and we introduced the ADNI dataset. We started with the history of machine learning and deep learning algorithms. Then, we provided the necessary information about the machine learning and deep learning algorithms that are utilized in this thesis. In order to understand the proposed pooling method in this thesis, the voxel-based morphometric analysis is explained. The patients profiles of the ADNI dataset

and the selected subjects for this thesis were introduced. Finally, the state of art in computational techniques in neuroimaging is provided.

In this thesis, we utilized convolutional neural networks combined with an autoencoder for unsupervised feature learning from neuroimaging data. In order to assess the classification power of the acquired features, we employed support vector machines. For comparison purposes, we also employed multilayer perceptrons and regression classifiers. As it is mentioned in this chapter, autoencoders and restricted Boltzmann machines are two similar approaches in deep learning. In this thesis, we have chosen autoencoders for its simplicity. In the following chapter, the methodology of the CNN model that is introduced in this thesis will be explained in detail.

3. METHODOLOGY

This chapter describes the research methodology used in this thesis. In order to be more specific, we followed quantitative research methodology approach to probe out the issue regarding how computer-aided diagnosis systems assist radiologists in the diagnosis and prognosis of dementias. Firstly, we refined our problem statements. Then, we searched the literature to develop our hypotheses. Next, we constructed our research plan including data collection, data preparation, computer-aided diagnosis (CAD) development and assessment.

3.1. PROBLEM STATEMENT

Several computer-aided diagnosis studies were proposed in the literature for dementias and cognitive disorders based on magnetic resonance imaging. Recent studies proposed computer-aided diagnosis systems to differentiate Alzheimer's disease (AD) patients from cognitively normal elderly (HC) with a high accuracies of 94.5% in the study of Magnin *et al.* [32], 96% in the study of Klöppel *et al.* [33], and 89.3% in the study of Vemuri *et al.* [34]. However, building a computer-aided diagnosis system for differentiating Alzheimer's disease from mild cognitive impairment (MCI) becomes a challenging and difficult task compared with the above works; because, cortical atrophy is a non-specific finding in Alzheimer's disease, and also, a patient with mild cognitive impairment may have similar atrophic patterns with AD patients.

Although mild cognitive impairment and Alzheimer's disease are distinct disorders, neurological researches showed that some of MCI patients eventually develop AD [102]. Because of the relation between them it has been of great interest to predict the risk of developing AD for MCI patients. Since AD shares many clinical symptoms with MCI [103], the existence of reliable biomarkers plays an important role in the identification of MCI patients who will develop AD. Those kind of MCI patients, called as "MCI-due-to-AD", are considered as in the symptomatic preface of Alzheimer's disease.

Previous researches have defined several biomarkers based on neuroimaging (i.e., structural magnetic resonance imaging (MRI), functional MRI, fluorodeoxyglucose – positron

emission tomography (FDG-PET) imaging) [104], cerebrospinal fluid (CSF) protein test [105] and blood test [106]. Thus, several neuroimaging modalities are being used to develop biomarkers for dementias, but it should be preferably considered the use of non-invasive ones [107]. Because, ethical issues may arise when considering the use of an invasive method in the diagnosis and treatment of a patient with an impaired mental function [108]. This fact oriented some researchers towards studies aiming to discover and use less invasive biomarkers for patients having dementias.

The blood-based biomarkers are considered as a minimally invasive option compared to the CSF-based biomarkers having a higher risk for patients [109]. However, the structural MRI scanning is a non-invasive method for acquiring structural images representing the morphology of brain structures in a high resolution [110]. By consequence, the structural MRI scanning became the most studied modality for dementias.

As an effort to investigate the possibility of a system allowing to make such a differentiation, in this thesis, we focused on the problem of predicting the prognostic type of a patients with MCI. This prognosis system would be based on the structural MRI of the patients, which is one of the most challenging issue in the field of computer-aided diagnosis of dementias.

To sum up, our first problem statement for this thesis work was mainly based on the recent growing research interest that is developing neuroimaging biomarkers based on non-invasive structural MRI data for identifying "MCI due to AD" patients.

Furthermore, in this thesis, we asked the question that "is it possible to predict the abnormal decline in cognitive abilities that is not related to normal aging in an early stage?". Nowadays, the assessment of ARCD is based on neuropsychological tools. The neuropsychological assessment is executed by several neuropsychological test to measure the cognitive decline of an individual in specific domains like memory, execution, orientation and attention. It is essential to predict the abnormal decline in cognitive abilities that is not related to normal aging in an early stage. The difficulties in the clinical interpretation of neurocognitive data for differential diagnosis of dementias have raised the use of computational techniques. Thus, our second problem statement of this thesis has arised as to discover the possibility of distinguishing the patients with a diagnosis of ARCD

from the patients with other type of dementias or cognitive disorders based on only neuropsychological assessment data.

3.2. RELATED WORKS

In order to narrow down our problem statements into specific hypotheses, we did comprehensive literature search in order to find out the studies that have the same diagnostic or prognostic goals with us and the studies that are utilized the deep learning techniques to extract features of MRI data.

Firstly, we did a literature search to figure out the studies that aimed to distinguish age-related cognitive decline (ARCD) from other dementias. There were no specific study that aimed to identify ARCD among other dementias in the literature. So, we expanded our search to discover related works that utilized neuropsychological test batteries for the classification purposes of dementias. Several computational techniques were employed in the domain of CAD models based on neuropsychological assessment data. Järvelin *et al.* applied machine learning methods to model aphasia related problems based on the naming test scores [76], [77]. In some other studies, the test scores of neurocognitive data were used to classify primary progressive aphasia subtypes [111] or to distinguish semantic dementia patients [112]. Utilizing random forest tree classifiers was recommended by Clark *et al.* to predict cognitive decline in the elderly having a greater risk of developing Alzheimer's diseases [113]. Although multilayer perceptrons were employed to distinguish normal elderly people from people presenting mild cognitive impairment, dementia, Parkinson's disease, stroke and traumatic brain injury [75]. Previous studies applied machine learning techniques successfully based on a single neuropsychological test score [114] or compared different machine learning algorithms like linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests [115]. Silva *et al.* [116] selected the better subset among four different verbal memory tests (logical memory, California verbal learning test, verbal paired-associate learning and digit span) to obtain the highest predictive power. According to the literature search, to the best of our knowledge, there were no study of computer-aided diagnosis systems based on neuropsychological assessment data exists to distinguish age-related cognitive decline from other dementias.

A second literature search was conducted to identify the studies that utilized computational techniques for predicting the risk of developing Alzheimer's disease for MCI patients as it is summarized in Table 3.1.

Table 3.1. Recent studies predicting the risk of developing AD for MCI

Studies	Longitudinal	Invasive	ADNI	Data type/ Number of Patients	ACC	TPR	SPC
Devanand et al. [84]	No	No	No	MRI+Clinical+Cognition 27 C-MCI, 79 NC-MCI	94.9%	92.6%	85.2%
Green et al. [117]	No	Yes	No	EEG+CSF 19 C-MCI, 17 NC-MCI	94.4%	94.7%	94.1%
Nazeri et al. [118]	Yes	Min	Yes	MRI+Plasma 110 C-MCI, 190 NC-MCI	93.5%	N/A	N/A
Douaut et al. [119]	No	Yes	No	MRI(DTI)+CSF 13 C-MCI, 22 NC-MCI	91%	85%	96%
Minhas et al. [83]	Yes	No	Yes	MRI+Cognition N/A	89.7%	87.5%	92.3%
Peters et al. [87]	No	No	No	MRI+Cognition 18 C-MCI, 22 NC-MCI	87.5%	90.9%	83.3%
Dukart et al. [120]	No	Min	Yes	MRI+PET+Genetic 177 C-MCI, 265 NC-MCI	86.8%	87.5%	77.8%
Ortiz et al. [121]	No	No	Yes	MRI+PET N/A	83%	67%	95%
Ardekani et al. [86]	Yes	Min	Yes	MRI+Cognition+Genetic 86 C-MCI, 78 NC-MCI	82.3%	86%	78.2%
Cheng et al.	No	Yes	Yes	MRI+PET+CSF 43 C-MCI, 56 NC-MCI	80.1%	85.3%	73.3%
Korolev et al. [122]	No	Yes	Yes	MRI+Clinical+CSF 139 C-MCI, 120 NC-MCI	80%	83%	76%
Willette et al. [85]	No	No	Yes	MRI+Cognition 76 C-MCI, 86 NC-MCI	80%	78.3%	81.5%
Suk et al. [70]	No	No	Yes	MRI+PET 76 C-MCI, 128 NC-MCI	75.9%	48%	95.2%
Suk et al. [70]	No	No	Yes	MRI 76 C-MCI, 128 NC-MCI	72.4%	36.7%	91%

Among the works listed in Table 3.1, the different kinds of data were utilized e.g. MRI, clinical information (Clinical), neuropsychological test scores (Cognition), genetic test results (Genetic), Electroencephalography (EEG), blood plasma test results (Plasma), CSF

test results and PET scans. The highest accuracy was achieved by combining cross-sectional structural MRI, clinic information and neurocognitive test scores with an accuracy of 94.9% [84]. In the work of Ardekani *et al.* [86], a measure of the hippocampal volumetric integrity (HVI) from a baseline and one-year follow-up structural MRI was utilized together with cognitive tests, genetics and demographic information. Minhas *et al.* [83] used the data derived from longitudinal MRI together with neuropsychological test score. In another work, Nazeri *et al.* [118] utilized plasma protein measurements together with tensor based morphometry (TBM) of MRI data to detect changes over time, which collecting blood plasma samples is considered as a minimally invasive method. Devanand *et al.* [84], Minhas *et al.* [83], Peters *et al.* [87], Ardekani *et al.* [86], Willette *et al.* [85] were utilized the neuropsychological assessment data in addition to the neuroimaging data, which may be difficult to perform for some of the patients with mild cognitive impairment.

Table 3.2. Recent studies utilized deep learning on neuroimaging

Studies	*Goal*	*Deep Learning Method*	*Data type/ Number of Patients*	*ACC*
Suk *et al.* [71]	C-MCI vs. NC-MCI	Deep SAE	MRI+PET+CSF	75.8%
Suk *et al.* [123]	C-MCI vs. NC-MCI	Deep SAE	MRI+PET+CSF	77.9%
Li *et al.* [124]	C-MCI vs. NC-MCI	Patch-based 3D-CNN	MRI+PET	72.4%
Suk *et al.* [70]	C-MCI vs. NC-MCI	Deep RBM	MRI	72.4%
Suk *et al.* [70]	C-MCI vs. NC-MCI	Deep RBM	MRI+PET	75.9%
Gupta *et al.* [81], [82]	AD vs. MCI	Patch-based 2D-CNN	MRI	86.3%
Payan *et al.* [92]	AD vs. MCI	Patch-based 3D-CNN	MRI	86.8%
Plis *et al.* [125]	SC vs. HC	Deep BN + RBM	fMRI	90%

Thirdly, the methodological motivation of this thesis, as it is described in section 1.1.2, led us to investigate the recent neuroimaging studies utilizing deep learning algorithms as in Table 3.2. The studies of Suk *et al.* [70], [71], [123] and Li *et al.*[124] aimed to classify MCI patients as likely convert (C-MCI) or remain stable (NC-MCI). Additionally, the diagnostic goal of the studies of Gupta *et al.* [81], [82] and Payan *et at.* [92] was to distinguish the patients with a diagnosis of AD from MCI. Finally, the schizophrenia (SC) patients were successfully distinguished from the healthy control (HC) individuals with an accuracy of 90% based on functional MRI (fMRI) data [125]. Of those, five studies were used deep

learning approaches to reduce the high-dimensionality of neuroimaging data; and only four of them were utilized deep learning to extract features directly from the images [81], [82], [92], [124]. Among those patch-based CNN studies, the studies of Gupta *et al.* [81], [82] and Payan *et al.* [92] were learnt the convolutional filters of CNN model by using an autoencoder. On the other hand, Li *et al.* [124] were trained CNN using back-propagation algorithm to learn the convolutional filters, which is very time and resource consuming operation. Suk *et al.* were utilized deep stacked-autoencoder (SAE) structures for dimensionality reduction purposes resulted in an accuracy of 75.8% [71] and 77.9% [123], which the MRI data is represented as average values of each atlas-defined region-of-interests (ROI) on GM tissue probability maps.

To sum up, the studies of Gupta *et al.* [81], [82], Payan *et al.* [92] and Plis *et al.* [125] were utilized deep learning approaches for the feature representation purposes while converting MRI data into quantitative data. Since the method applied by Plis *et al.* [125] is highly resource consuming for high-dimensional data and our data is much more bigger than their data, we decided to use the patch-based convolutional neural networks.

Furthermore, the study of Suk *et al.* was selected as a reference study for performance comparison due to sharing the same predictive goal with us and utilizing only structural MRI data, which yielded an accuracy of 72.4% [70].

3.3. HYPOTHESIS

In this thesis, after searching the literature comprehensively, we stated our hypotheses related to the problem statements. In other words, we narrowed down our problem statements into five hypotheses in order to achieve objectives of this thesis.

Two of five hypotheses was related to our first problem statement that aims to distinguish age-related cognitive decline from other dementias. The first hypothesis (hypothesis-I) says that age-related cognitive decline can be distinguished from other dementias using machine learning algorithms based on neuropsychological assessment data. The second hypothesis (hypothesis-II) was relevant to the hypothesis-I that a feature selection method can improve the performance of the machine learning algorithms.

Other three hypotheses were connected with the second problem statement searching the possibility of predicting the risk of developing Alzheimer's disease for MCI patients. The third hypothesis (hypothesis-III) says that the deep learning models based on cross-sectional structural MRI data provide biomarkers indicating the risk of developing AD for MCI patients. The forth hypothesis (hypothesis-IV) is similar to the third hypothesis. It claims that the predictive power of longitudinal structural MRI data is higher than cross-sectional data for deep leaning approaches. Finally, the last hypothesis (hypothesis-V) says that a pooling method for CNN based on the significant topographic regions of the brain can produce features with a high prognostic power.

3.4. DATASETS

In this thesis, we obtained three different datasets that are the neuropsychological assessment data of age-related cognitive decline and dementias, the cross-sectional and the longitudinal structural MRI data of MCI patients.

3.4.1. Neuropsychological Assessment Data

In order to evaluate the hypothesis-I and hypothesis-II, we acquired the dataset retrospectively from the neurology department of the Maltepe University Hospital, which contains neuropsychological measures collected from patients who are diagnosed with one among the following diseases: age-related cognitive decline, mild cognitive impairment, vascular dementia or Alzheimer's disease. A total of 125 participants were in our databank. Of these, 19 cases were excluded due to methodological issues related with the completion of all neuropsychological tests and significant variation of age and education variables compared to average values of the groups.

Of the remaining patients (n=106; mean age = 71.97 ± 7.37 years), 30 patients were diagnosed with ARCD (age = 70.07 ± 7.90 years, education = 10.57 ± 3.97 years, 22 female, 8 male), 20 were diagnosed with probable Alzheimer's disease (age = 75.45 ± 5.53 years, education = 9.75 ± 4.30 years, 11 female, 9 male), 21 were diagnosed with probable vascular dementia (age = 72.09 ± 6.61 years, education = 9.14 ± 3.59 years, 13 female, 8 male) and 35 were diagnosed with amnestic mild cognitive impairment (age = 71.54 ± 7.94 - years,

education = 10.34 ± 3.68 years, 18 female, 17 male). Each subject was assigned to a clinical dementia rating scale (CDR) [126] in which a value of zero indicates no dementia and values of 0.5, 1, 2 and 3 indicate very mild, mild, moderate and severe dementia, respectively. Alzheimer's disease and vascular dementia cases, who were classified by clinicians with a CDR value of 0.5 or 1, were selected for this thesis.

All ARCD patients met the diagnostic criteria described by the National Institute of Mental Health work group [9] and the criteria of the International Psychogeriatric Association Working Party. The first criterion, the subjective impairment, is a report demonstrating that cognitive function has declined and the second one, the objective impairment, shows the difficulties in any of these cognitive tests such as memory and learning, attention and concentration, abstract thinking (problem solving, abstraction), language and visuo-spatial functioning in at least one standard deviation measurement (SD) below the age and education norms. Psychometric tests have normal values according to age and education level [127]. The probable AD patients were diagnosed according to the National Institute of Neurological and Communicative Disorders and Stroke and by the Alzheimer's Disease and Related Disorders Association [128]. In addition, the diagnostic criterias for probable VD were set according to the National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherche et l'Enseignement en Neurosciences criteria [18]. Finally, the MCI patients were diagnosed according to the Albert *et al.* [129] criteria.

In this dataset, all subjects were examined individually with a neurocognitive test battery (NCTB) that aimed to assess the functions of orientation, attention-executive functions, language, visuo-spatial skills and memory by the neurology department of the Maltepe University Hospital. The measures of cognitive functions of the subjects included the Turkish versions of the tests. For the assessment of orientation, scores of personal and current information and the Benton judgment of line orientation test [130], were used. Attention-executive functions were examined with the following tests: digit span forward and digit span backward subscales of the Wechsler memory scale-revised (WMS-R), mental control subscale of the Wechsler memory scale-revised, verbal fluency tests (that include naming animals for semantic fluency and K, A, S letters for phonemic fluency), the similarities subtest of the Wechsler adult intelligence scale-revised (WAIS-R), interpretation of Turkish proverbs (three proverbs, with the rating of zero to three points each), the Turkish

version of word fluency tests, a clock drawing test (a zero-to-three point scoring system was used) [131] and the interference score, incorrect response score and spontaneous correction scores of the Stroop test [132]. To assess language function, the Boston naming test (BNT) was performed [133]. Verbal fluency tests (both semantic and phonemic) were classified in language functions in addition to the attention-executive functions. Visuo-spatial skills were examined with the Benton face recognition test, the Benton judgment of line orientation test (visuo-spatial perception) and a copying drawings test (three simple and one complex drawings, with the rating of one point each) [130]. Both visual and verbal memory were assessed. Visual memory was examined with the visual reproduction subtest of the WMS-R including immediate recall, delayed recall and recognition scores. Verbal memory was examined with the Öktem-VMPT (Ö-VMPT) test including scores of immediate memory, total learning, the highest learning, delayed recall and recognition [134]. WMS-R logical memory subscale is also used for verbal memory assessment.

3.4.2. Neuroimaging Data

In order to evaluate the hypotheses III, IV and V, the structural MRI data of mild cognitive impairment patients were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) in the preprocessed forms of T1-weighted MR images (3 Tesla) acquired with a magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence with data access permission from the ADNI. We selected two datasets for this thesis: cross-sectional dataset and longitudinal dataset as explained in the next subsections in more detail.

3.4.2.1. Cross-sectional MRI

A total of 305 patients with a baseline diagnosis of mild cognitive impairment were taken as the subjects of this thesis to test the hypothesis III and V, whose written informed consents were obtained by the ADNI before acquiring.

The subjects were separated into two groups: those who developed Alzheimer's disease during their follow-up (n=140; C-MCI) and those who remained stable in the mild cognitive impairment state (n=165; NC-MCI). Among those 140 C-MCI subjects, 26 were diagnosed

as having Alzheimer's disease six months after the baseline assessment, 38 after 12 months, 11 after 18 months, 39 after 24 months, 18 after 36 months, five after 48 months, one after 72 months, and two after 96 months.

3.4.2.2. *Longitudinal MRI*

Furthermore, a total of 168 ADNI patients with a baseline diagnosis of MCI having both baseline and 12-months follow-up structural magnetic resonance imaging scans were taken as subjects to evaluate the hypothesis IV and V. Additionally, these patients have at least four continuous follow-ups that the time interval between two follow-up is six months or one-year.

Among them, 73 patients who developed Alzheimer's disease were grouped as Converted-MCI (C-MCI) and 95 patients who remained stable in mild cognitive impairment were grouped as non-Converted-MCI (NC-MCI). Among the 73 C-MCI group subjects, 11 of them were diagnosed with AD 18 months later, 37 of them after 24 months, 17 of them after 36 months, five of them after 48 months, one of them after 72 months and two of them after 96 months from the baseline examination.

The pre-processed forms of the selected baseline and one-year follow-up 3-Tesla T1-w MRI scans were downloaded from the ADNI data archive.

3.5. STATISTICAL ANALYSIS

Several statistical tests were employed in this thesis for comparing and presenting diagnostic groups in demographic profiles and for comparing predictive performances of machine learning algorithms. Statistical data analysis are conducted using MATLAB (Matlab and Statistics Toolbox Release 7.3, The MathWorks, Inc., Natick, Massachusetts, USA).

3.5.1. Describing Data

Descriptive statistics were used to present the demographic-clinic profiles or neuropsychological test scores of the diagnostic groups. Among these variables, categorical

variables (e.g. gender) were given as the number and percentage of female and male groups in corresponding groups. Further, continuous variables (e.g. age, education, mini-mental state examination or neuropsychological test scores) are expressed with mean and standard deviation.

3.5.2. Comparing Diagnostic Groups

Demographic-clinic profiles or neuropsychological test scores among more than two groups were compared by a Kruskal-Wallis statistical test [135], which is a non-parametric alternative to the one-way analysis of variance (ANOVA) [136]. The ANOVA assumes that the dependent variable is normally distributed.

In order to identify significant difference between two groups on a variable; a two-tailed Wilcoxon rank sum test [137] and a Pearson's Chi-square test were employed for corresponding continuous variables (e.g. age, education and mini-mental state examination) and categorical variables (e.g. gender), respectively. An acquired p-value smaller than the selected alpha level indicates that there is a statistically significant difference between two groups on the tested variable.

The Wilcoxon rank-sum test is the non-parametric version of the two-sample t-test with a null hypothesis saying that the two groups comes from the same distribution with the same median [138].

A Wilcoxon rank-sum test is executed as follows: All observations are sorted from smallest to largest, which the rank of an observation is determined by its sort order. Then, ranks of each groups are summed up, separately. Next, a U statistics of the first group is calculated as follows:

$$U = n_1 * n_2 + \left[\frac{n_1 * (n_1 + 1)}{2} \right] - R_1, \tag{3.1}$$

where n_1 and n_2 are the number of observations in each groups, separately, and R_1 is the total rank of the first groups.

Then, z-score is calculated as:

$$z = \frac{U - ((n_1 * n_2)/2)}{\sqrt{(n_1 * n_2 * (n_1 + n_2 + 1))/12}} \qquad (3.2)$$

Finally, a critical z-value is calculated for a given significance level alpha to interpret the ontained results. Such that, a calculated z-value greater than the critical z-value rejects the null hypothesis states that there is a statistically significant difference between medians of two groups.

Additionally, the Kruskal-Wallis test [135] is a non-parametric alternative to the one-way analysis of variance (ANOVA) [136], where the data is replaced by its ranks. Then, the H-value of Kruskal-Wallis test is calculated for k number of groups as follows:

$$H = \left[12/(n * (n + 1))\right] * \left[\sum_{i=1}^{k} \frac{R_i^2}{N_i}\right], \qquad (3.3)$$

where n is the total number of observations, N_i is the number of observations in i^{th} groups and R_i is the sum of the ranks of all observations in i^{th} groups. Finally, a critical H-value is calculated to interprete the acquired H-value. In this thesis, k refers to the number of groups, n refers to the total number of patients in all groups and N_i is the number of patients in the ith group.

3.5.3. Comparing Classifier Performances

In this thesis, a Cochran's Q test was used to assess the significant differences between the responses of more than two classification algorithms on the same dataset [139]. Furthermore, McNemar's tests were implemented for multiple post hoc performance comparisons between pairs of classifiers. The Cochran's Q test is utilized to test the statistically significantly different outcome acquired from more than two treatments (in our case, the machine learning classifiers) for the same individuals. The Cochran's Q test applied on two treatments is equivalent to the McNemar's test [140].

The Q statistics is calculated for k number of treatments as follows:

$$Q = \frac{(k-1) * [(k * \sum G^2) - (\sum G)^2]}{(k * \sum L) - \sum L^2}, \tag{3.4}$$

where G_i indicates the sum of the values of ith treatments for all observations and L_i indicates the sum of the values of ith observation for all treatments.

Finally, a critical χ^2 value is calculated with a degree of freedom $(k-1)$ to compare with the obtained Q-value. A Q-value greater than the critical χ^2 value rejects the null hypothesis indicating that those treatments has significantly different effect on treatments. In this thesis, k refers to the number of classification algorithms to compare their performances that is bigger than two, and the observation refers to the prediction outcome of each algorithms.

3.6. METHODS
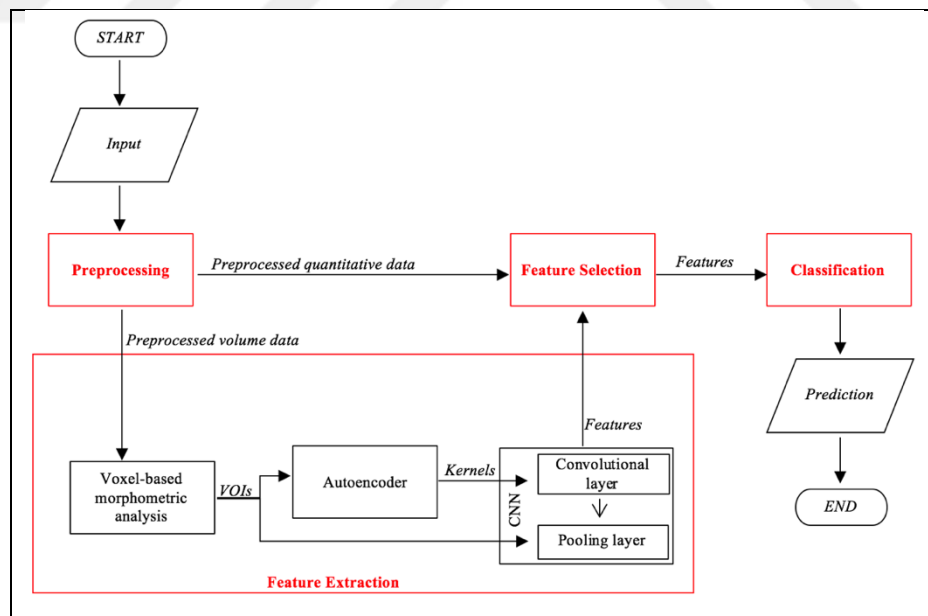
Figure 3.1. Flowchart for developing CAD systems.

In this thesis, computer-aided diagnosis systems for dementias were built in order to test our hypothesis. These CAD systems consisted of four common main stages: preprocessing, feature extraction, feature selection and classification as it is shown in Figure 3.1. In

preprocessing stage, input data is converted to a format that algorithms work with. The feature extraction stage only applies to volumetric data. In the feature selection stage, we tackled with the high-dimensionality of the extracted features. Finally we assessed the predictive power of the acquired features in classification stage. The following subsections descibes each of those main stages, separately, in detail.

3.6.1. Data Preprocessing

As it is explained in Section 3.4, the data types used in this thesis are the neuropsychological assessment data and the MRI data. The neuropsychological assessment data is a quantitative data type that is composed of the neurocognitive test scores of the patients with age-related cognitive decline or other dementias. On the other hand, the MRI data is the structural MR scans of the patients with MCI who developed AD or remained stable in MCI state. The acquired MRI scans are three-dimensional volumes that are composed of a set of axial images. Preprocessing methods are different for each type of data as they explained in the following subsections.

3.6.1.1. *Neuropsychological Assessment Data*

In this thesis, all neurocognitive test scores of a patient is considered as a feature vector of the patient, so all those feature vectors are used to train machine learning algorithms as it is. Our neuropsychological test battery was including 26 different tests, so the dimension of feature vector of a particular patient was 26 in this thesis.

3.6.1.2. *Cross-sectional MRI Data*

For the cross-sectional data, the normalization and segmentation steps of VBM were performed using the "VBM8" extension of the SPM8 toolbox in Matlab 2010a (The MathWorks, Inc., Natick, Massachusetts, USA) to acquire grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) tissue probability maps of the aligned structural images. The spatially normalized and modulated GM tissue probability maps were smoothed with a Full-Width Half-Maximum (FWHM) of 8 mm using SPM8, which the acquired volume is called as ***preprocessed volume data*** as in Figure 3.1.

3.6.1.3. Longitudinal MRI Data

In this thesis, the longitudinal structural MRI data of a patient is composed of MR volumes acquired in two different time points: the baseline and the 12-months follow-up. Those images are processed by a longitudinal analysis with a novel Jacobian-based method, which is proposed and implemented in this thesis.

This novel Jacobian-based method takes the differences between two MRI volumes and superposes them into the follow-up MRI as an attempt to create a single volume, called as ***preprocessed volume data*** as in Figure 3.1 that will include the current state and also mark the anatomical parts that will be deteriorated or altered.

This longitudinal pre-processing phase comprises of three steps: the creation of two grey matter volumes from the MRI scans belonging to the baseline and the follow-up examination, the creation of an enhanced follow-up volume that incorporates the differences between the two volumes and a smoothing operation of the resulted volume. As a first step, the baseline and the 12-months follow-up grey matter volumes were created from the MRI patients' images using the '*Segment Longitudinal Data*' function of the CAT. Those grey matter volumes were used in the second step. As a second step, we compute a grey matter volume of longitudinal changes for each patient using their baseline and 12-months follow-up exam.

In summary, we attempt to detect the differences in grey matter between the baseline and the 12-months follow-up. After detecting these differences, we add them to the follow-up volume. The result is an "enhanced" grey matter volume where an eventually progression in atrophy will be represented together with the actual state of the brain. In other words, we attempt to accentuate the eventual differences that a brain can have in time as a painter that will accentuate parts of a drawing with its pencil.

The complete procedure is described in mathematically as follows: Let a transformation function $f : R^3 \rightarrow R^3$, where R^3 is the three-dimensional space and $f = \begin{Bmatrix} b \\ v \end{Bmatrix}$, where b is the baseline and v is the 12-months follow-up grey matter volumes of a patient.

Again, let $f_i : R^2 \rightarrow R^2$ where R^2 is the two-dimensional space and $f_i = \begin{cases} b_i \\ v_i \end{cases}$, where b_i and v_i are the ith slices of the b and v volumes respectively. Then, we define the Jacobian matrix of f_i in x and y direction with the derivatives of b_i and v_i in each directions as:

$$J_i = \begin{bmatrix} \dfrac{\partial b_i}{\partial x} & \dfrac{\partial b_i}{\partial y} \\ \dfrac{\partial v_i}{\partial x} & \dfrac{\partial v_i}{\partial y} \end{bmatrix}. \tag{3.5}$$

Then, the determinant value of the Jacobian matrix is defined as:

$$\det(J_i) = \frac{\partial b_i}{\partial x} \cdot \frac{\partial v_i}{\partial y} - \frac{\partial b_i}{\partial y} \cdot \frac{\partial v_i}{\partial x}. \tag{3.6}$$

Finally, the absolute value of the Jacobian determinant is calculated as:

$$Z_i = abs\ (\det(J_i)) \tag{3.7}$$

If we apply this procedure for every slice, it will result in the volume Z. This acquired volume represents all differences that exist between corresponding slices of the baseline and the 12-months follow-up MRI scans. We decided to take the absolute value of the Jacobian determinant because we are interested of every alternation that occurs in grey matter between the consecutive exams. Moreover, our experimental results showed that the use of the absolute value in the Jacobian determinant significantly increased the accuracy of the system.

Later on, we acquired the enhanced follow-up volumes V by adding the 12-months follow-up volume f together with the volume Z which incorporates the differences of the two exams:

$$V = f + Z, \tag{3.8}$$

where V is the volume of the enhanced follow-up.

Figure 3.2 presents this procedure for the generation of a particular slice of the enhanced follow-up patient's exam: the baseline and the corresponding 12-months follow-up slice are used to generate the Jacobian determinant in absolute values.

Then, as shown in Figure 3.2, an element-wise addition with the original follow-up slice is performed in order to create the final enhanced follow-up slice. This procedure is repeated for all slices and the enhanced follow-up volume is generated. Thirdly and finally, the enhanced follow-up volumes were smoothed with a full-width half-maximum (FWHM) of 8 mm using SPM, which were used to feed the input layer of the CAD system, which is called as ***preprocessed volume data*** as in Figure 3.1.
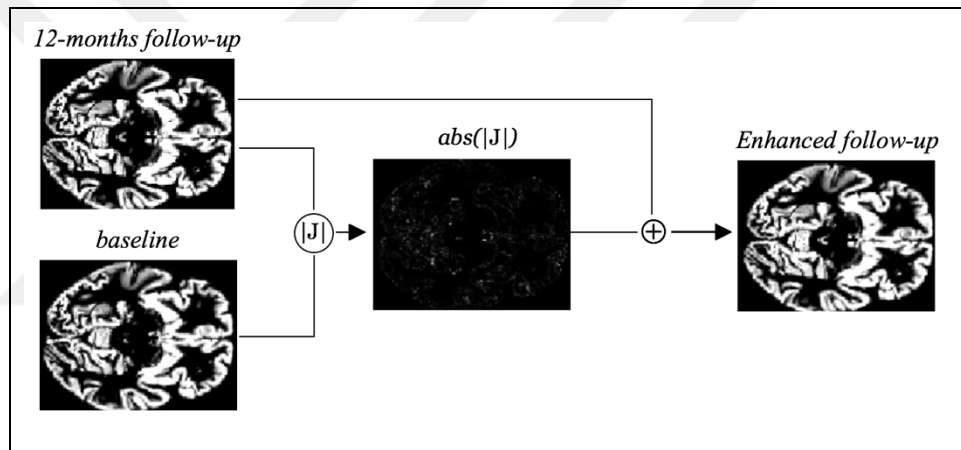
Figure 3.2. Flowchart for preprocessing longitudinal data

3.6.2. Feature Extraction

The section aims to produce a set of features from the preprocessed volume data that are capable of predicting whether a particular patient with mild cognitive impairment will remain stable in the diagnosis of mild cognitive impairment or eventually will develop Alzheimer's disease.

In this thesis, the developed a novel feature extraction system comprises of a voxel-based morphometric analysis, an autoencoder and a convolutional neural network, as it is seen in Figure 3.1. The following subsections describes the novel single-layer convolutional neural network schema.

3.6.2.1. *Voxel-based Morphometry*

In the statistical analysis stage of VBM analysis of ***the preprocessed volume data*** was executed for convolutional neural network to create SPM-t or SPM-F contrast maps. In order to create contrast maps, we utilized the SPM software, which calculates *t*-value or *F*-value for every voxel in the GM tissue based on the general linear model (GLM). The GLM is a statistical tool to express a dependent variable as a function of a weighted sum of independent variables, as:

$$y = X\beta + c, \tag{3.9}$$

where y is the dependent variable, X is the independent variable, β is the estimated parameters and c is the residuals. The matrix X is called as design matrix in SPM, which each row represents a patient and the columns stores the other related variables.

The parameters β are estimated by ordinary least squares algorithm, which can be expressed as in terms of X and y, as:

$$\hat{\beta} = (X^T X)^{-1} X^T y. \tag{3.10}$$

Thus, the residual error r is the difference between the real y and the estimated $\hat{y}$; and estimated error variance $\hat{\sigma}^2$ is defined as:

$$\hat{\sigma}^2 = ((r\char94 T \, r))/df, \tag{3.11}$$

where the degrees of freedom df is calculated as the difference between the number of observations and the rank of the matrix X.

Finally, the *t*-statistics and *F*-statistics are calculated using the linear combination of the estimated parameters $\hat{\beta}$, as $c^T \hat{\beta}$. The volume where the t-value from t-statistics or F-value from F-statistics belonging to each voxel is displayed, is called SPM-t contrast map or SPM-F contrast map, respectively.

The selection of the significance level value for the SPM-t contrast map or SPM-F contrast map has repercussions for type-I and type-II errors [141]. Type-I errors occur when a high value significance is selected, which leads to a low t-value threshold thus including more voxels and increasing by consequence the false positive rate. Type-II errors occur when a low value of significant level (α) is chosen, which results in a high t-value threshold that excludes more voxels and increases by consequence the false negative rate. After a series of experiments, an appropriate statistical significance level (α) was selected .

3.6.2.2. *Autoencoder*

This section describes the concept of autoencoders and how it is employed in our novel CNN model to extract features from MRI data. An autoencoder is a feed-forward artificial neural network aiming to learn the latent representation of the data [58]. An autoencoder neural network applies back propagation algorithm for training the network, such that the output data is almost same as the input data. In this network, the number of hidden neurons can be smaller or bigger than the number of input neurons. When the number of hidden neurons is smaller than the number of input neurons, the network learns the compressed representation of the data. Otherwise, it learns the sparse representation of the data.
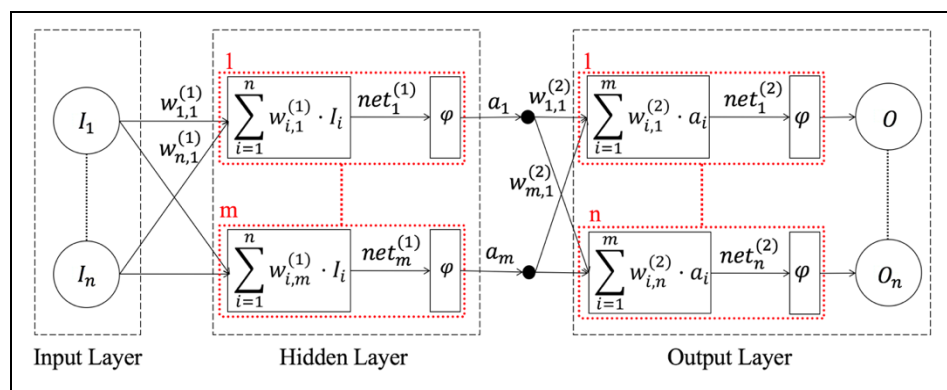
Figure 3.3. Architecture of an autoencoder

The general schema of an autoencoder with a single hidden layer is shown in Figure 3.3. In this example autoencoder network, the input I and the output O are the n-dimensional vectorised data. Each element of input and output is represented with subscript. The autoencoder is trained to satisfy a condition where the input and the output will be almost

same. In this figure, the hidden layer is composed of m hidden neurons. Each element of the input vector I is connected to each hidden neuron with the weight matrix $w^{(1)}$ of $n \times m$ size, where the its value stands for the weight of the connection between i^{th} input element and j^{th} hidden neuron.

As a result, the generated vector a is the latent representation of the input data I. The output value a_j of the jth hidden neuron is calculated as in Equation (3.12):

$$a_j = \varphi\left(\sum_{i=1}^{n} w_{i,j}^{(1)} \cdot I_i\right), \qquad (3.12)$$

where φ is the activation function and the input of the activation function $net_j^{(1)}$ is the weighted sum of the values of the elements of the input vector I. The operation is repeated for all hidden neurons. Thus, the acquired vector a is the latent representation of the input data I. In other words, an autoencoder can be formulated as:

$$O = decoder\big(encoder(I)\big), \qquad (3.13)$$

where the encoder is the function that calculates the vector a and the decoder is the function that reconstructs the input data vector from the vector a. This is how an autoencoder used as a dimensionality reduction purposes.

However, in this thesis, we didn't use that property of the autoencoder in this thesis. We utilized autoencoders to learn convolutional filters. In order to do this, the Equation (3.12) is rewritten to cover all elements of the vector a, then each column of the weight matrix is called as basis vector (see Equation (3.14)).

$$[a_1 \quad \dots \quad a_m] = \varphi\left([I_1 \quad \dots \quad I_n]\begin{bmatrix} w_{1,1}^{(1)} & \dots & w_{1,m}^{(1)} \\ \vdots & \vdots & \vdots \\ w_{n,1}^{(1)} & \dots & w_{n,m}^{(1)} \end{bmatrix}\right) \qquad (3.14)$$

In the case where the activation function φ is the linear identity function, the encoder part of the autoencoder corresponds with the principal component analysis method, but with a

difference that the basis vectors of the encoder part of the autoencoder are non-orthogonal unlike principal component analysis. Finally, each column of the weight matrix is one of the learnt filter to be used as filter in the convolutional layer of convolutional neural network models.

In addition to these, although the backpropagation learning algorithm is a supervised algorithm, due to the use of the data itself as a label for training the autoencoder, the training of an autoencoder can be considered as an unsupervised learning procedure.

To sum up, the purpose of using an autoencoder in this thesis is that generating convolutional filters for CNN model. For this purpose, a set of patches are extracted from the significant regions of the brain. These filters can be considered as an abstraction or a general representation of the information included in the regions of interest. During the training process, we used a set of patches from the preprocessed volume of all patients and we generated a set of filters. Thus, every filter can be considered as an image in the same size with the patches and its pixels correspond to the weights of a particular hidden layer. Extracted patches were vectorised to feed the input layer of the autoencoder.

After training, the weights of each hidden unit compose a filter. As a result, a number of filters of the same number as the hidden units are obtained.

3.6.2.3. *Traditional Convolutional Neural Network*

Before continue to explain our novel CNN model, it is better to explain how a traditional convolutional neural network works. A CNN architecture can be formed in a variety of ways, but the simplest one is composed of a set of convolutional and subsampling/pooling layers and optionally a supervised classifier as a final layer.

Figure 3.4 illustrates a traditional structure of a two-layered convolutional neural network. In this configuration, the input layer contains an $m \times n$ image that represents an instance. Firstly, the input image is convolved with $k^{(1)}$ filters denoted by $C_n^{(1)}$, where $n \in \{1, 2, .., k^{(1)}\}$. Thus, the convolving operation of the first layer produces $k^{(1)}$ feature maps with the same size as the input image.
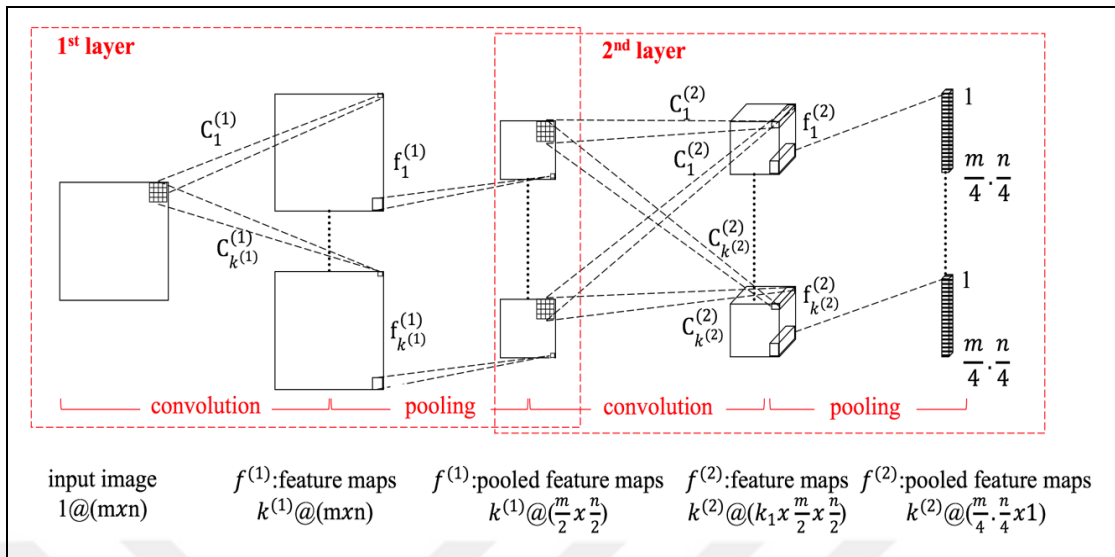
Figure 3.4. Architecture of a two-layered convolutional neural network

Now, it is better to explain how a feature map is convolved with a filter for a better understanding of convolutional neural network. The convolution is a mathematical operation on two functions that indicates the amount of overlap of the two functions as sliding the filter over the first function, which plays an essential role in extracting important features from the data in signal and image processing.

In a more abstract definition, the convolution is the integral of the point-wise multiplication of the two functions f and g that is given by:

$$(f * g)(t) = \int f(\tau)\, g(t - \tau) d\tau , \tag{3.15}$$

where f and g are continuous function of time τ and t is a time-offset. In image processing context, the two-dimensional image convolution for a particular (x, y) position is mathematically represented as:

$$(f * g)(x, y) = \sum_{v=y-kh}^{y+kh} \sum_{u=x-kw}^{x+kw} f(u, v) g(x - u, y - v) , \tag{3.16}$$

where f is the input and g is the filter with a size of $(2 \cdot kw - 1) \times (2 \cdot kh - 1)$.

In order to have a better understanding of the convolution, the calculation of the function $(f * g)(x, y)$ with an 7×7 input image and a 3×3 filter (named as kernel also) at the point $x = 2$ and $y = 2$ is illustrated in Figure 3.5.
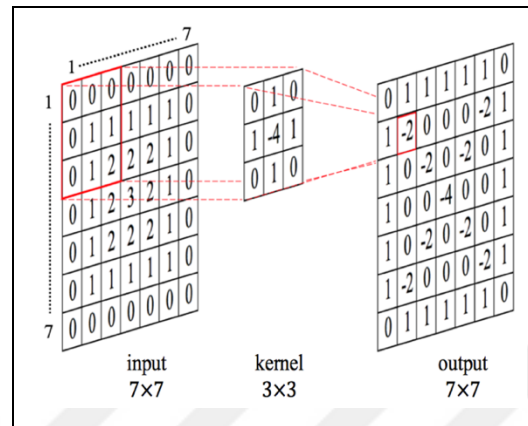
Figure 3.5. Illustration of a convolution operation of an image with a laplacian filter

Therefore, according to the Equation (3.16), the value of that point is calculated as $(f * g)(2, 2) = 0 \cdot 0 + 0 \cdot 1 + 0 \cdot 0 + 0 \cdot 1 + 1 \cdot (-4) + 1 \cdot 1 + 0 \cdot 0 + 1 \cdot 1 + 2 \cdot 0$, which the result is "-2". In a convolution operation, each filter extracts different features out of the data. For example, the laplacian filter is used for calculating the second derivative of an image as it is illustrated in Figure 3.5.

After convolutional layer, each of the acquired feature maps is pooled separately to create feature vectors. The aim of the following pooling operation is to reduce the size of the feature maps. In that case, the size of feature maps were reduced by 0.5 using 2×2 pooling areas, which the acquired pooled feature maps form the input of the second layer. The filters of the second layer is denoted by $C_n^{(2)}$, where $n \in \{1, 2, .., k^{(2)}\}$ in the Figure 3.4. The number and size of kernels may vary for each layer. Since more than one image stands for an instance in the second layer, the convolution results are aggregated for each filter separately that compose $k^{(2)}$ features maps in the size of $k^{(1)} \times (m/2) \times (n/2)$.

Finally, the pooled feature maps are vectorised in the size of $1 \times (k^{(2)} \cdot (m/4) \cdot (m/4))$ to be used as feature vectors in the classification tasks. In this thesis, we assessed the predictive power of the acquired feature vectors of patients using support vector machines.

3.6.2.4. Novel Convolutional Neural Network Model

In this thesis, a novel single-layer convolutional neural network model is introduced, which comprises an input, a convolutional layer, a pooling layer for extracting prognostic features from the structural magnetic resonance imaging data [142]. To be more specific, we introduced a novel single-layer convolutional neural network schema that employed an autoencoder for the convolutional layer and a SPM-t contrast map or a SPM-F contrast map used in the pooling layer to define the pooling clusters. In addition, the contrast maps were also indirectly used in the training procedure of an autoencoder by identifying appropriate voxels for defining where the patches will be sampled. Finally, the acquired prognostic features is used later as an input for the traditional machine learning classifier for assessment. This section is organized as follows. Firstly, we introduce the creating of contrast map. Then, the training of autoencoders with the help of contrast maps, and then, the purpose of using an autoencoder in this novel CNN model are presented. Next, the meaning of contrast maps in the pooling layer is explained. Finally, we introduced the novel CNN model utilizing the autoencoder and VBM.

First of all, two voxel-based morphometric analyses were performed on preprocessed cross-sectional volumes and preprocessed longitudinal volumes, separately, using the statistical parametric mapping (SPM) software. To be more specific in terms of this thesis, a SPM-t contrast map was built using the preprocessed cross-sectional volumes of two groups of patients (C-MCI vs. NC-MCI) by performing a voxel-wise two-sample t-test with a one-sided alternative hypothesis stating that the patients with C-MCI group shows more grey matter atrophy in their baseline then NC-MCI group. Similarly, a two-sample t-test was designed on the preprocessed longitudinal volumes to specify statistically significantly different voxel positions between the two groups, which is resulted in a SPM-F contrast map. The significance level was selected as 0.005 with an extend threshold of 600 for both contrast maps in this thesis.

Secondly, the autoencoders were trained with the help of contrast maps. The purpose of using autoencoders in this CNN model is to create the filters to be used in the convolutional layer. To be more specific in terms of this thesis, we trained two autoencoders using the preprocessed cross-sectional volumes and preprocessed longitudinal volumes, separately.

The first auto-decoder was trained by a set of patches of size 11×11 that are composed by 1000 patches sampled per preprocessed cross-sectional volumes. The patches are sampled from the region determined by the SPM-t contrast map. The number of hidden units of the autoencoder was determined after a series of extensive tests implying different numbers of hidden units based on the best success rate of the system. The autoencoder with 10 hidden units was trained according to the vectorised representations of these patches with a linearly decreasing learning rate going from 0.9 to 0.25, a stable learning scale of 0.9, a momentum of 0.2, and L2-weight regularization penalty of 0.0001 over a total of 20 epochs. Additionally, the second autoencoder was trained by a set of patches of size 15×15 with the help of SPM-F contrast map. During the training process, we used 457.800 patches from the preprocessed longitudinal data of all patients and we generated 12 filters as described in Figure 3.6. Extracted patches were vectorised to feed the input layer of the autoencoder as in the figure. After training, the weights of each hidden unit composed a filter in size of 15×15. Since the system has 12 hidden units, we come up with 12 filters. Those learnt filters are further used in the convolutional layer of the novel CAD models.
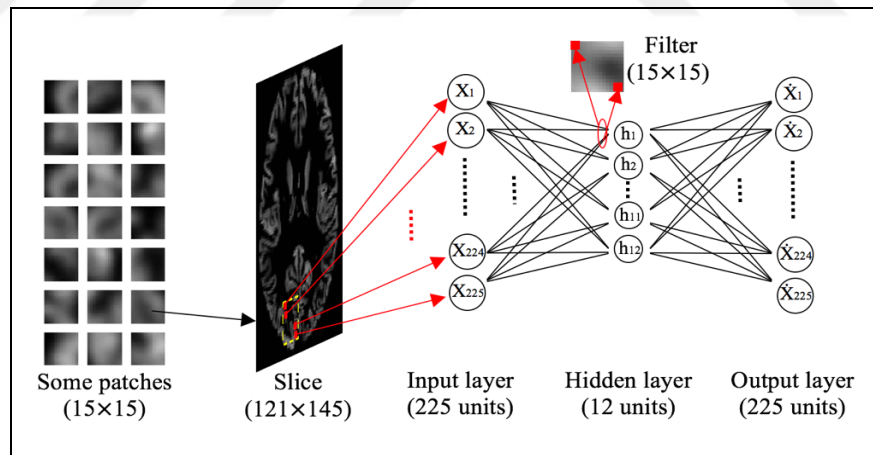
Figure 3.6. Flowchart for generating filters using an autoencoder

Thirdly, the main purpose of using the contrast maps in novel CNN model is to developed a novel pooling schema to test the hypothesis-V in this thesis.

In other words, the contrast maps were used to determine which voxels will be pooled together, named as clusters. Depends on the selected significance level, we came up with a set of clusters.
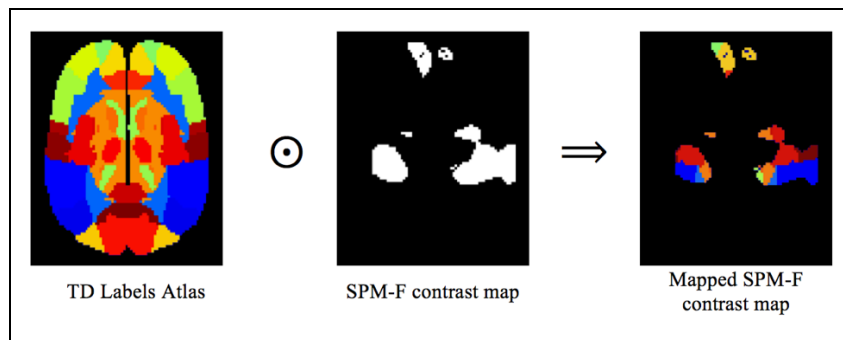
Figure 3.7. Flowchart for defining pooling clusters

For determining pooling regions, an atlas-based region-of-interest (ROI) analysis was performed on the clusters using the WFU PickAtlas software for mapping the regions of interest according to the "TD Labels" atlas as in Figure 3.7. In the analyses related with cross-sectional data, this method couldn't be applicable. So, the clusters were used as it is determined by SPM-t contrast maps for the cross-sectional data. In other words, the step explained in Figure 3.7 is skipped for the cross-sectional experiments.
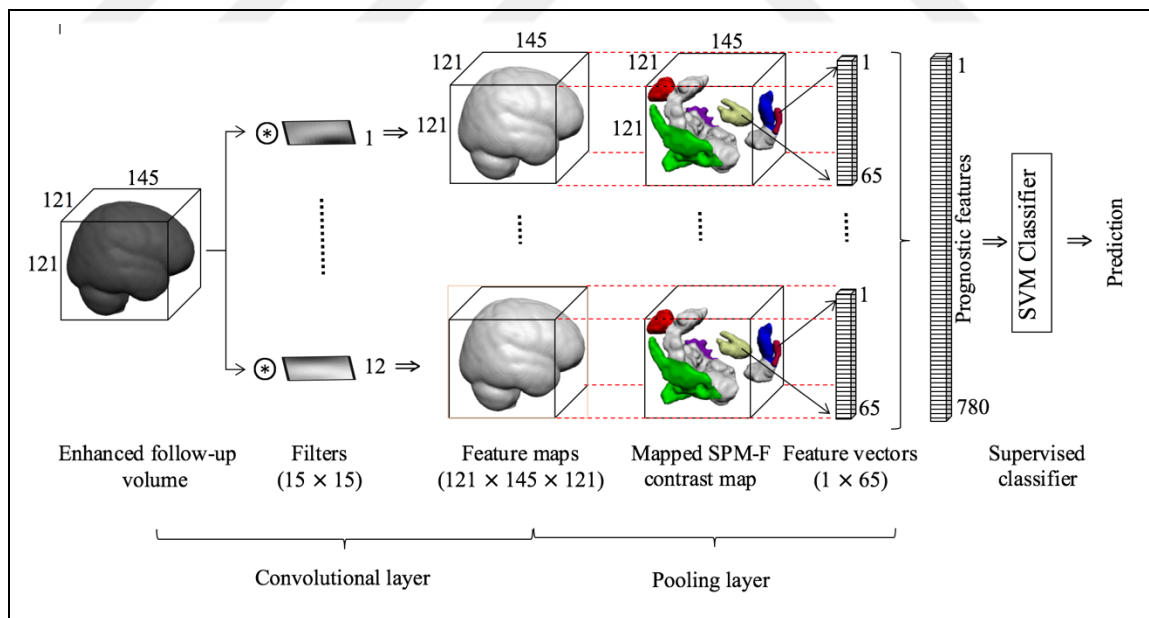
Figure 3.8. Architecture of the novel convolutional neural network for longitudinal data

Finally, after acquiring the filters by an autoencoder and the pooling regions by a contrast map, we are ready to built the novel CNN model for extracting prognostic features. Figure 3.8 presents the steps that are needed for creating the prognostic feature vector of a patient

explained through the preprocessed longitudinal data as an example. After several experiments for the longitudinal data, we achieved the highest results with 12 filters in a size of 15×15 and 65 clusters caused by a significance level of 0.005 with and extend threshold of 600. After the acquisition of the patients' preprocessed longitudinal volumes (enhanced follow-up volumes), we apply the following convolution procedure: for every patient, its brain volume is convolved with each of the 12 filters calculated before. As mentioned above, these filters may be considered as an abstraction of the characteristics indicating if an AD conversion will happen or not. For a particular brain volume, the convolution is performed with each slide of the volume. In total, 12 new brain volumes are generated (named as "*feature maps*"). Then, for each of them, using the mapped SPM-F contrast map, 65 volumes of interest are determined. These regions, as explained before, indicate the parts of individual brain anatomical regions that are considered statistically important after performing the VBM analysis. The next step is to calculate the mean value of each volume of interest. In total, 65 values are generated for each feature map named as "*feature vector*". As we have 12 convolved brain volumes (or 12 feature maps), in the end, a total of 780 values are produced named "*prognostic features*" for longitudinal experiments.

Similarly, another single-layer convolutional neural network model was constructed using the preprocessed cross-sectional volumes in this thesis as in Figure 3.8. The best results were achieved when we trained an autoencoder with 10 hidden units and we selected the significance level as 0.005 and extend threshold of 600. In that case, we utilized the clusters as they defined by VBM analysis. Thus, we had five clusters that resulted in 50 prognostic features per patients for cross-sectional experiments.

3.6.3. Feature Selection

In order to evalute the hypothesis-II, a Whitney ranksum test [137] was employed to compare all possible pairs of treatment groups on each neurcognitive test scores with a significance level of 0.01.

Then, a subset of the neurocognitive test scores was selected for each possible pair of diagnostic groups based on these Mann-Whitney test statistics in order to train machine learning classifier [143].

Moreover, we implemented a support vector machine based recursive feature elimination method (SVM-RFE) [144] in order to select the most significant features among the prognostic features acquired from the longitudinal MRI data due to its high-dimensionality. This method is based on a feature scoring procedure indicating the contribution of each feature to the construction of the SVM hyperplane during the training of the SVM classifier. For this purpose, we used the SVM-RFE implementation of Weka.

3.6.4. Classification

Finally, the acquired features are processed in a traditional machine learning classifier for assessing their prognosis powers. In this thesis, we utilized two common machine learning algorithms to assess our obtained feature vectors that are multilayer perceptrons and support vector machine classifiers. The following subsection describes those two algorithms.

3.6.4.1. *Multilayer Perceptrons*

In this thesis, we have utilized the multilayer perceptrons to build the classification models based on neuropsychological test scores. Thus, it is good to provide some theoretical information about the multilayer perceptrons.
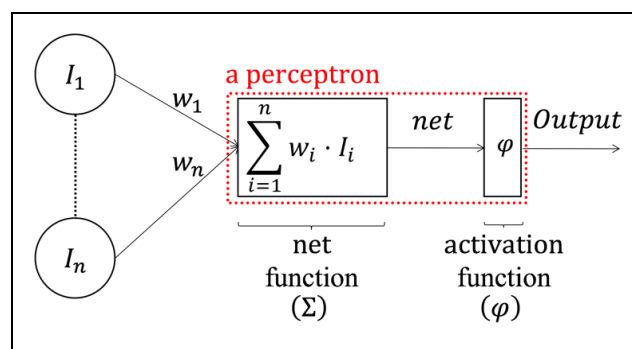
Figure 3.9. Architecture of a perceptron

The building blocks of a multilayer percepton are the perceptrons that are discovered by Frank Rosenblatt [50]. The perceptron is a single unit, which comprises a net input function and an activation function in a serial order as it is illustrated in Figure 3.9. The net input function calculates the weighted sum of the input values. The activation function can be any

differentiable function that calculates the output of that perceptron according to its net value calculated by the net function of it. Generally, a linear function, a sigmoid function or a logistic function is used as an activation function to transforms the net value. Each perceptron in a multilayered structure may have different kind of activation function.

Thus, as the name implies, the multilayer percepton is the finite directed acyclic graph of perceptrons. In this network structure, the output of a perceptron is used as input by the other perceptrons that are connected next to it. A perceptron may have more than one input connection but have only one output connection. If a perceptron has no input connection then it is called as "input perceptron", similarly, if it has no output connection then it is called as "output perceptron". In other words, a multilayer perceptron is a collection of perceptrons organized in a layered structures, where the first layer is the input layer, the last layer is the output layer and remainings are named as the hidden layers.

The connection between the perceptrons in a multilayer perceptron network has a scalar value, named as "weight", which the value of the weight shows the strength of the bond between the perceptrons. Generally, the weights of a multilayer perceptron are trained by the backpropagation algorithm [145]. The algorithms back propagates the calculated error from the output layer to the input layer and then performs gradient descent in order to minimize error.

The common definition of an error in backpropagation is the square of the Euclidean distance between the input vector I and the output vector O, as:

$$E = \frac{1}{2b} \sum_{x=1}^{b} \left\| I^{(x)} - O^{(x)} \right\|^2 \tag{3.17}$$

where b is the batch size that is the number of input data forward propagated together and each individual data vector is indicated by a superscript indexes one through b.

Then, the partial derivative of the total output error E with respect to one of the weights of the output layer gives the amount of change occurred in the total output error when a change occurred in that weight.

By applying the chain rule, the derivative of the total output error E with respect to the weight $w_{i,j}$ becomes as:

$$\frac{\partial E}{\partial w_{i,j}} = \frac{\partial net_j}{\partial w_{i,j}} \cdot \frac{\partial O_j}{\partial net_j} \cdot \frac{\partial E}{\partial O_j},$$

(3.18)

where the weight of the connection between the ith perceptron in the hidden layer prior to the output layer and the jth perceptrons in the output layer is indicated by $w_{i,j}$, E is the total output error, and finally, the symbol net_j and the symbol O_j stands for the net value and output value of the jth perceptrons in the output layer, respectively.
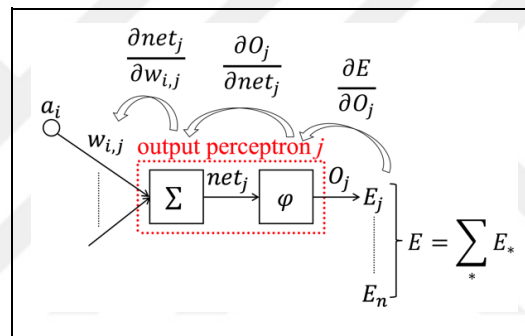
Figure 3.10. Illustration of the chain rule

For a better understanding, the illustration of the chain rule can be seen in Figure 3.10, where the symbol a_i indicates the net value of the ith perceptron in the hidden layer prior to the output layer and each individual error is indicated by a subscript.

As we said before, the function φ can be any differentiable function, but in general, the sigmoid function is used. If we assume that the activation function φ is the sigmoid function, then the derivative of it is becomes as: $O_j(1 - O_j)$. This is why a differentiable function is needed as the activation function.

When we put derivatives of the all parts in Equation (3.18), the following is obtained:

$$\frac{\partial E}{\partial w_{i,j}} = a_i \cdot \left(O_j \cdot (1 - O_j) \right) \cdot (O_j - I_j).$$

(3.19)

Finally, the new value of the $w_{i,j}$ is calculated as:

$$\widehat{w}_{i,j} = w_{i,j} - \eta \cdot \frac{\partial E}{\partial w_{i,j}}, \qquad (3.20)$$

where η is a coefficient called as *learning rate*. Finally, this procedure is repeated for all weights of the hidden layer and the backpropagation is applied for the prior layer using the output vector a of the hidden layer. After training the weights of the network, it will be ready for predicting the output of a new input data.

In this thesis, we constructed one MLP classifiers per each pairs of groups to assess the predictive power of neuropsychological test scores of patients with an input layer of size 26 (one for each test score) and an output layer of size two (one for each groups). Additionally, to test the hypothesis-II, we constructed similar MLP classifiers having a difference in the input layer. Such that, the input layer is fed by the selected subset of test scores.

3.6.4.2. *Support Vector Machine*

In this thesis, we utilized support vector machines to build the classification models for neuropsychological test scores and also for assessing the predictive power of the gained neuroimaging features. The support vector machine, as it is mentioned before, is proposed by Vapnik *et al.* as a result of the synergy occurred between machine learning and statistics in 1950s [51]. A classic support vector machine is suitable for linearly separable datasets. With the help of kernel trick, it is possible to separate the datasets that are non-separable.

The aim of the training of a support vector machine is to find the most suitable hyperplane that classifies the dataset clearly among all possible hyperplane. The training procedure of a SVM is to learn a function $f(x)$ with a given set of labelled training data (x_i, y_i) as:

$$f(x) = w^T x + b, \qquad (3.21)$$

where w is the weight matrix, x_i is a given input data in the n-dimensional feature space, $y_i \in \{-1, 1\}$ indicates the class membership of x_i and b is the bias.

This learnt function $f(x)$ is used for the prediction of class membership of a new data x in the same feature space as:

$$y = \begin{cases} -1, & f(x) < 0 \\ 1, & f(x) \geq 0 \end{cases}, \tag{3.22}$$

where y is the predicted class membership for x. In this thesis, x is an $m{\times}n$ dimensional matrix, which m refers to the number of subjects and n refers to the number of features acquired from either neuropychnological data or MRI data. Furthermore, the class memberships y refers to the diagnostic groups.

3.6.5. Evaluation

Classification performances were measured in terms of accuracy (ACC), sensitivity (TPR), specificity (SPC), precision (PPV) and F-measure (F1).

$$Accuracy = (TP + TN)/(TP + FN + FP + TN), \tag{3.23}$$

$$Sensitivity = TP/(TP + FN), \tag{3.24}$$

$$Specificity = TN/(FP + TN), \tag{3.25}$$

$$Precision = TP/(TP + FP), \tag{3.26}$$

$$F - measure = 2TP/(2TP + FN + FP) \tag{3.27}$$

The accuracy is defined as the percentage of all correct identifications among the whole groups (in Eq. 3.23). The sensitivity measures the percentage of correctly identified patients within the primary group (in Eq. 3.24), and similarly, the specificity measures the percentage of correctly identified patients among the secondary group patients (in Eq. 3.25). Furthermore, the precision shows the percentage of correctly identified the primary group patients to the number of patients predicted as in the same group (in Eq. 3.26). Finally, the F-measure is the harmonic mean of precision and sensitivity as it is formulated (in Eq. 3.27).

Furthermore, in the above equations, the *TP* (true positives) and the *TN* (true negatives) are the number of correctly classified the primary and the secondary patients respectively. Finally, the *FP* (false positives) and the *FN* (false negatives) denotes the numbers of incorrectly classified patients.

In this thesis, a ten-fold cross-validation procedure was applied to the machine learning classifiers [146]. With this procedure the data set is divided into ten parts and for each repetition of training 90% of the dataset is the training set and the remaining 10% constitutes the test part.

Only for the neuropsychological assessment data, the *ClassifierSubsetEval* attribute evaluator of WEKA [147] software was applied to every diagnostic pairs with the selected features using a ''Best First'' ranking scheme for each classifier model, separately, to calculate the ''merits'' of the features that reflect the classifier's specific characteristics.

3.7. SUMMARY OF THE CHAPTER

This chapter presented the essential parts of the scientific research methodology of this thesis. A quantitative research methodology approach was followed in this thesis. We defined two problem statements: The first problem statement was about the issue of predicting the risk of developing AD for the MCI patients based on non-invasive data. The second problem statement focused on the problem of differentiating age-related cognitive decline from other dementias using only neuropsychological assessment data. After a comprehensive literature search, the problem statements were narrowed down to five hypotheses, which two of them were related with the first problem statement and remaining ones were related with the second problem statement. The hypothesis-I says that neuropsychological assessment data is sufficient to build a CAD model for distinguishing ARCD from dementias and the hypothesis-II says that a feature selection method will improve the performance of the CAD models. Similarly, the hypothesis-III and hypothesis-IV were constructed to test the predictive power of the cross-sectional and longitudinal MRI data, respectively, in determining the risk of developing AD for MCI patients. Finally, the hypothesis-V was claiming that considering significant topographic regions of the brain in the pooling stage of the CNNs will improve the performance of the models.

In order to test our hypothesis, we acquired the dataset retrospectively from the neurology department of the Maltepe University Hospital, which contains neuropsychological measures collected from patients who are diagnosed with one among the following diseases: age-related cognitive decline, MCI, vascular dementia or Alzheimer's disease. Furthermore, the baseline and follow-up MRIs of the MCI patients in the format of preprocessed forms of T1-weighted MR images (3 Tesla) with a magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) with data access permission of ADNI.

Finally, the diagnostic or prognostic features were extracted from either neuropsychological test scores or MRI data. A novel single-layer CNN model were introduced to extract features of structural MRI data. Next, traditional machine learning algorithms were employed to assess the predictive performances of the acquired features are given in this section.

4. RESULTS

This chapter describes the findings of the experiemental evaluations executed to test the hypotheses of this thesis; including, descriptive and inferential statistics of the diagnostic groups and datasets, the results of the voxel-based morphometric analysis, the presentation of the extracted features and the evaluation of the predictive performances of the developed CAD models.

4.1. MATERIAL

We obtained the three datasets as previously introduced in Section 3.4. To recall it, the first dataset was composed of the neuropsychological test scores of the patients with a diagnosis of ARCD, MCI, AD or VD. The baseline structural MR images of the MCI patients who remained stable in MCI or developed AD formed the second dataset. Finally, both baseline and 12-months follow-up structural MR images of the same kind of MCI patients were obtained as the third dataset.

4.1.1. Subject

Table 4.1. Descriptive statistics of the groups having neuropsychological data.

	ARCD (n=30)	MCI (n=35)	AD (n=20)	VD (n=21)	p-value
Gender (F/M)	22/8	18/17	11/9	13/8	=0.32
Age	70.1±7.9	71.5±7.9	75.5±5.5	72.1±6.6	=0.06
Education	10.6±3.9	10.3±3.7	9.8±4.3	9.1±3.6	=0.57

Table 4.1. shows the demographic profiles of the patient groups in the first dataset on gender (female (F)/ male (M)), age (in years) and education (in years). The p-values from Kruskal-Wallis test statistics on each demographic variable is given, where a p-value smaller than the selected significance level of 0.01 indicates that there is no significant difference among the four diagnostic groups on corresponding variable.

The descriptive statistics and statistical comparison of the patients groups of the second dataset in gender, age, education (in years) and mini-mental state examination (MMSE) score at baseline examination are shown in Table 4.2. The statistical analysis showed that the two groups did not differ significantly with regards gender, age, and education ($p > 0.001$). On the other hand, the MMSE scores of patients with NC-MCI (the group of MCI patients who remained stable in MCI) were significantly higher than in patients with C-MCI (the group of MCI patients who developed AD after baseline) with a significance level of 0.001.

Table 4.2. Descriptive statistics of the groups having cross-sectional MRI data

	NC-MCI	C-MCI	Statistics
Gender	75 (45.5 %) / 90 (54.5 %)	54 (38.6 %) / 86 (61.4 %)	p=0.2253
Age	70.89±7.26; [56-89]	73.12±7.26; [55-87]	p=0.003
Education	16.42±2.55; [11-20]	15.91±2.81; [6-20]	p=0.134
MMSE	29.02±0.80; [28-30]	27.06±1.84; [23-30]	p<0.001

The descriptive statistics and statistical comparison of the patient groups of the third dataset on gender, age, education and MMSE score at baseline (MMSE-bl) and 12-months follow-up (MMSE-fl) are shown in Table 4.3. The statistical analysis showed that the two groups did not differ significantly in gender, age and education at the significance level of 0.001. On the other hand, the median of MMSE scores of the NC-MCI patients were significantly higher than those of the C-MCI patients ($p < 0.001$).

Table 4.3. Descriptive statistics of the groups having longitudinal MRI data

	NC-MCI (n=95)	C-MCI (n=73)	Statistics
Gender	46 (48.4 %) / 49 (51.6 %)	25 (34.2 %) / 48 (65.7 %)	p=0.065
Age	70.56±7.05; [57-89]	73.75±7.33; [56-87]	p=0.002
Education	16.42±2.59; [11-20]	16.29±2.74; [11-20]	p=0.776
MMSE-bl	29.04±0.79; [28-30]	27.30±1.66; [24-30]	p<0.001
MMSE-fl	28.62±1.45; [24-30]	26.33±1.81; [22-30]	p<0.001

As it is shown in Table 4.4, no significant difference in MMSE score between the baseline indicated by *MMSE-bl* and the 12-months follow-up indicated by *MMSE-fl* examination was

noted for the NC-MCI group (t-value: -2.85; df: 94; p-value = 0.003). On the other hand, the mean MMSE score at the 12-months follow-up examination of the converted mild cognitive impairment group was significantly reduced (t-value: -4.94; df: 72; p-value < 0.001). The group means of NC-MCI is 29.04 and 28.62 for baseline and follow-up, respectively. Similarly, the baseline and follow-up means are 27.30 and 26.33 respectively for C-MCI.

Table 4.4. The paired t-test statistics on the longitudinal MMSE scores

	Group means		Statistics
	MMSE-bl	*MMSE-fl*	
NC-MCI	29.04	28.62	t-value: -2.85; p-value = 0.003
C-MCI	27.30	26.33	t-value: -4.94; p-value < 0.001

4.1.2. Datasets

In this section, the descriptive and inferential statistics of the neuropsychological test scores of the patients groups obtained from the neurology department of Maltepe University Hospital are given. Furthermore, one of the baseline structural MR images of the patient with MCI is given with its corresponding grey matter tissue segment. Finally, the baseline and 12-months follow-up images of a particular patient is presented as an example to the longitudinal data.

4.1.2.1. Neuropsychological Assessment Data

Table 4.5. shows the mean and the standard deviation of neurocognitive test scores for each groups supported by Kruskal-Wallis test statistics. An asterisk (*) indicates a significance level of 0.05 and two asterisks (**) are used to denote the significance level as 0.01. In this thesis, the critical significance level of 0.01 is selected for the statistical analysis. At this selected significance level ($\alpha = 0.01$), it is shown in Table 4.5 that there is no statistically significant difference on WMS-R personal and current information (p=0.03), WMS-R orientation (p=0.014), WMS-R digit span forward (p=0.05), WMS-R mental control (p=0.02), Stroop spontaneous corrections (p=0.03), face recognition (p=0.08) and copying drawings (p=0.06).

Table 4.5. Descriptive and Kruskal-Wallis statistics on each neurocognitive test scores

		ARCD	MCI	AD	VD	p
Orientation						
Personal information		5.57±0.7	5.34±0.8	4.85±0.9	5.33±0.9	*
Orientation		4.97±0.2	4.69±0.6	4.35±1.0	4.76±0.4	*
Attention-executive functions						
Digit span forward		5.67±1.2	5.20±0.87	4.95±0.94	5.00±0.9	*
Digit span backward		4.47±1.2	3.66±0.7	3.30±0.7	3.86±1.8	**
Mental control		9.97±3.2	9.43±3.1	8.15±2.5	7.71±2.9	*
Semantic fluency		20.27±5.0	16.69±5.5	14.30±5.7	14.38±5.6	**
Fonemic fluency		33.67±15.3	25.11±11.4	23.00±13.3	18.86±8.1	**
Similarities		13.10±2.5	11.77±2.7	8.75±3.9	11.14±3.6	**
Proverbs		8.77±0.4	8.20±1.5	7.40±2.3	7.38±2.4	**
Clock drawing		2.63±0.7	2.60±0.7	2.00±0.9	2.00±1.0	**
Stroop interference		48.0±21.5	82.7±53.8	77.0±36.9	118.9±65.7	**
Stroop incorrect responses		1.03±1.9	7.09±11.5	7.20±11.6	5.29±7.8	**
Stroop spontaneous		2.53±2.9	2.86±2.2	3.65±2.7	5.14±4.1	*
Language						
Naming (BNT)		29.87±1.9	28.00±5.4	27.25±2.6	27.10±4.4	**
Semantic fluency		20.27±5.0	16.69±5.5	14.30±5.7	14.38±5.6	**
Fonemic fluency		33.6±15.3	25.1±11.4	23.0±13.3	18.86±8.1	**
Visuo-spatial functions						
Face recognition		41.83±3.1	41.20±4.4	39.40±2.6	40.19±3.8	0.08
Copying drawings		3.80±0.4	3.89±0.5	3.70±0.6	3.48±0.8	0.06
Memory functions						
WMS-R	*VR immediate recall*	11.37±2.1	9.57±2.3	7.85±2.5	8.33±2.7	**
	VR delayed recall	9.87±4.0	6.03±3.9	3.45±3.0	7.00±3.1	**
	VR recognition	3.13±1.2	2.34±1.4	1.40±1.2	2.48±1.3	**
	LM immediate recall	16.23±3.0	14.0±3.3	11.30±4.0	15.57±2.8	**
	LM delayed recall	16.07±3.7	12.31±4.3	9.80±5.3	15.57±3.0	**
Ö-VMP	*Immediate memory*	5.13±2.3	3.97±1.7	3.30±1.7	3.62±1.0	**
	Total learning score	103.8±24.6	72.0±18.5	61.2±13.3	74.24±17.5	**
	Highest learning score	13.33±2.4	10.23±2.3	8.60±1.5	10.76±2.1	**
	Delayed free recall	11.07±3.1	5.94±3.3	3.45±3.1	7.76±3.1	**
	Recognition score	15.00±0.0	13.63±1.5	11.75±2.8	14.05±1.4	**

4.1.2.2. *Cross-sectional MRI Data*

Figure 4.1 shows an axial slice of the 3T structural MR image acquired in the baseline examination of a 65 years-old woman with a diagnosis of MCI who subsequently progressed to AD 12 months after this baseline examination. Figure 4.1 (a) shows the 3T T1-weighted bias-corrected MR scan (170 sagittal slices, slice thickness=1.2 mm, matrix 256×256, magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence) as acquired from the ADNI dataset. Additionally, the corresponding preprocessed cross-sectional volume of the patient is shown in Figure 4.1 (b).
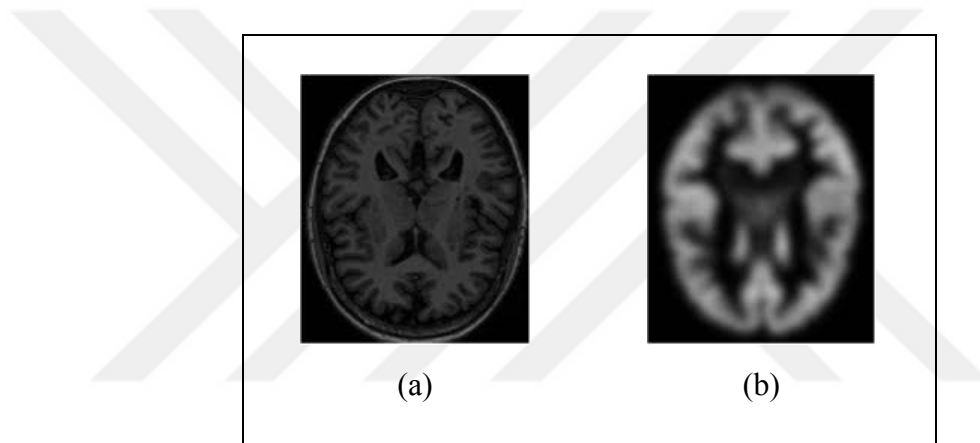
Figure 4.1. Cross-sectional MRI data of a patient. (a) Original baseline, (b) Preprocessed

4.1.2.3. *Longitudinal MRI Data*
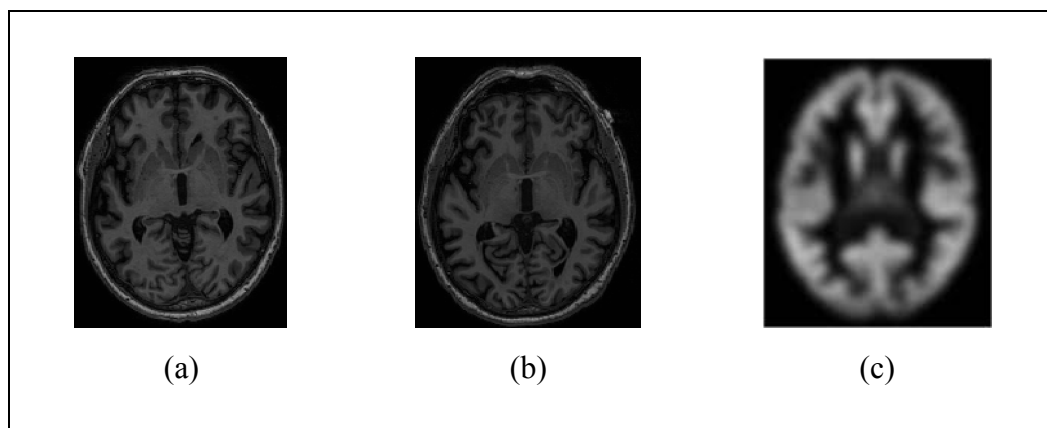
Figure 4.2. Longitudinal MRI data of a patient. (a) Original baseline, (a) Original 12-months follow-up, (c) Preprocessed

Figure 4.2 shows the 3T structural MR images acquired in the baseline and 12-months follow-up examination of a 74 years-old man with a diagnosis of MCI who subsequently progressed to AD 18 months after this baseline examination. Figure 4.2 (a) shows an axial slice of the 3T T1-weighted bias-corrected baseline MR scan (170 sagittal slices, slice thickness=1.2 mm, matrix 256×256, magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence) as acquired from the ADNI dataset. Similarly, Figure 4.2 (b) shows an axial slice of the 12-months follow-up scan of the same patient. Additionally, the corresponding preprocessed cross-sectional volume of the patient is shown in Figure 4.2 (c).

4.2. FEATURE EXTRACTION

In this thesis, the feature extraction procedure is subject to only the MRI data. The feature extraction procedure of MRI data is performed based on the novel convolutional neural network system explained in Section 3.6.2.5.

Several novel CNN models were constructed to create prognostic features with three main configuration parameters: those are the size of patches, the number of filters and the significance level for contrast maps. Thus, the six configurations were designed with the combination of one patch size (11x11), three different number of filters (10, 25, 50) and two different significance levels (0.001% and 0.005%) for the preprocessed cross-sectional volumes. For that case, we selected only one patch size (11x11), because other size of patches didn't worked well.

Similarly for the preprocessed longitudinal volumes, the combination of three different patch sizes (11x11, 13x13, 15x15) and four different number of filters (8, 10, 12, 14) resulted in 12 experiments for the evaluation of the prognostic power of the preprocessed longitudinal volumes.

The results based on these configurations are given in this section. Firstly, the contrast maps are shown with the descriptive information of the clusters obtained at the selected significance level. Then, the learnt convolutional filters are presented. Finally, an example feature map is given to a better understanding of the purpose of convolutional filters to highlight features of images.

4.2.1. Contrast Maps

As it is explained in Section 3.6.2.1, two voxel-based morphometric analyses were conducted that resulted in a SPM-t contrast map and a SPM-F contrast map for preprocessed cross-sectional and longitudinal volumes, respectively. This section shows these contrast maps and the acquired clusters at the selected significance level.

4.2.1.1. SPM-t Contrast Map

Figure 4.3 (a) shows the SPM-t contrast map of grey matter lower concentration differences observed between the C-MCI and NC-MCI groups with an extend threshold of 200 and a significance level of 0.005 based on the preprocessed cross-sectional volumes of the patients. The superimposition of the SPM-t contrast map based on the individual brain atlases using statistical parametric mapping of 116 predefined segments (IBASPM-116) atlas of the *WFU Pickatlas* software is presented in Figure 4.3 (b). The significantly reduced grey matter regions in the C-MCI group compared to NC-MCI group were thresholded at a family-wise error-corrected p-value of 0.005 (threshold t-value is 5.08). This significance level yielded five significant clusters. As an extra test, the same procedure was repeated at the family-wise error-corrected-corrected significance level of 0.001, which resulted in seven clusters.
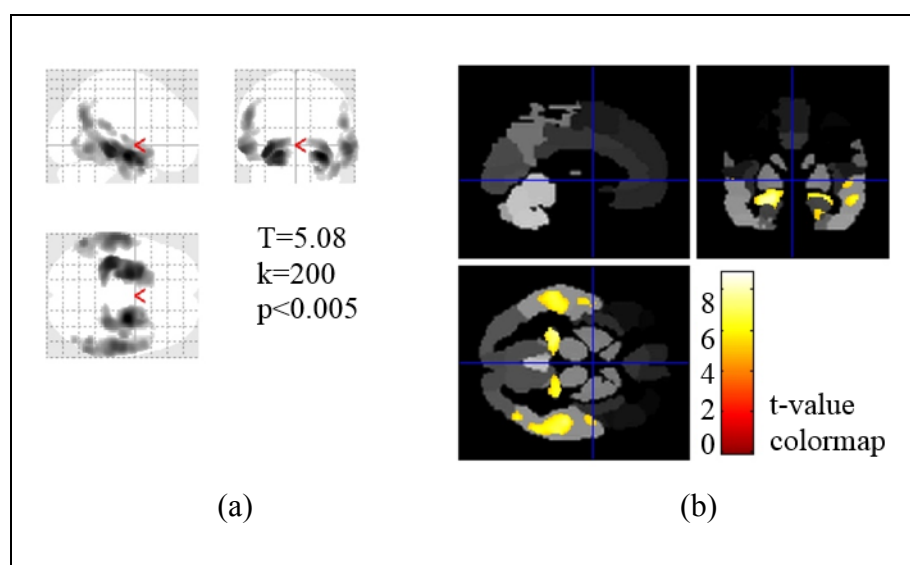
Figure 4.3. SPM-t contrast map based on the cross-sectional data. (a) SPM-t contrast map,
(b) Superimposed SPM-t contrast map on the atlas

The descriptive statistics of the five clusters that acquired at the significance level of 0.005 are shown in Table 4.6. Each row corresponds to a cluster. Such that, the *voxels* column shows the number of voxels within the cluster, the *t-value* column shows the mean and standard deviation of the t-values of the voxels within the cluster, the *maximum* column shows the maximum t-value observed in the cluster and the *IBASPM-116 atlas region* column gives the anatomic brain region of the voxel with the maximum t-value.

Table 4.6. Descriptive statistics of the five clusters (p<0.005)

	Voxels	t-value	Maximum	IBASPM-116 atlas region
1	3713	5.82±0.57	7.64	Right superior temporal gyrus
2	3109	6.45±0.87	8.39	Left amygdala
3	2921	6.04±0.75	8.85	Right hippocampus
4	2744	5.86±0.59	7.58	Left superior temporal gyrus
5	302	5.55±0.34	6.40	Left supramarginal gyrus

Similarly, the descriptive statistics of the seven clusters are shown in Table 4.7, which are obtained when the SPM-t contrast map is thresholded at the significance level of 0.001.

Table 4.7. Descriptive statistics of the seven clusters (p<0.001)

	Voxels	t-value	Maximum	IBASPM-116 atlas region
1	2628	6.67±0.77	8.39	Left amygdala
2	2114	6.32±0.69	8.85	Right superior temporal gyrus
3	2050	6.16±0.50	7.64	Right hippocampus
4	1762	6.16±0.53	7.58	Left middle temporal gyrus
5	245	5.90±0.33	6.71	Left angular
6	153	5.82±0.26	6.40	Right inferior temporal gyrus
7	110	5.63±0.12	5.87	Right superior temporal gyrus

4.2.1.2. SPM-F Contrast Map

Figure 4.4 shows the SPM-F contrast map of statistically significantly different voxels between two groups with an uncorrected p-value of 0.005 and an extend threshold of 600 created using the preprocessed longitudinal volumes. The SPM-F contrast map defines a

total of 63119 significant voxels that have a higher F-value than the selected threshold F-value. A total of 6961voxels of the SPM-F contrast map are located in the sub-gyral region of the anatomical ROI mask. Similarly, some other prevailing anatomical labelling are parahippocampa gyrus (n=7188), middle temporal gyrus (n=6501), superior temporal gyrus (n=7378) and extra-nuclear (n=3122).
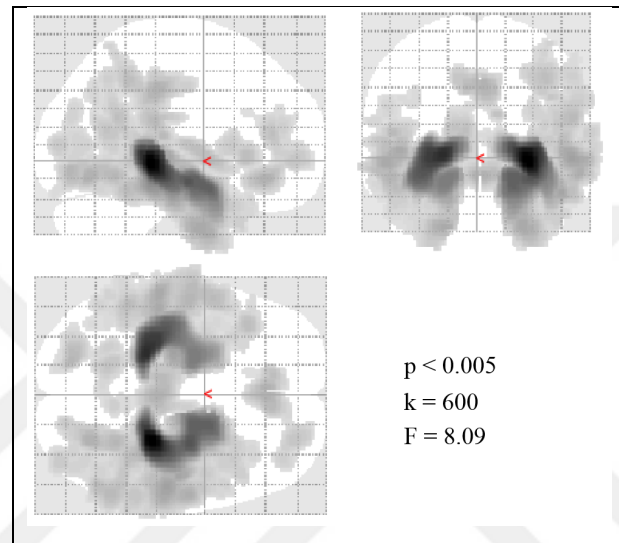
Figure 4.4. The SPM-F based on the longitudinal data

4.2.2. Convolutional Filters
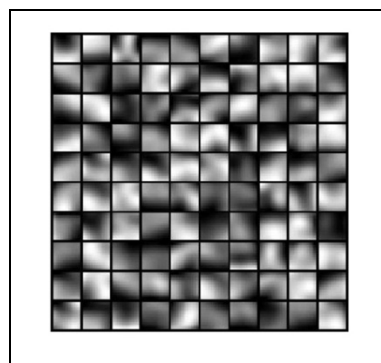
Figure 4.5. Some patches sampled from the preprocessed cross-sectional volumes

Figure 4.5 shows some of the patches in size of 11that sampled from the preprocessed cross-sectional volumes. These patches were used to train an autoencoder for learning the convolutional filters.

The convolutional filters learnt by an autoencoder with 10 hidden units are shown in Figure 4.6, where each subimage corresponds to a single convolutional filter with a size of 11×11. Those convolutional filters are sensitive to detect prognostic features of the preprocessed cross-sectional volumes of the patients with a diagnosis of MCI.
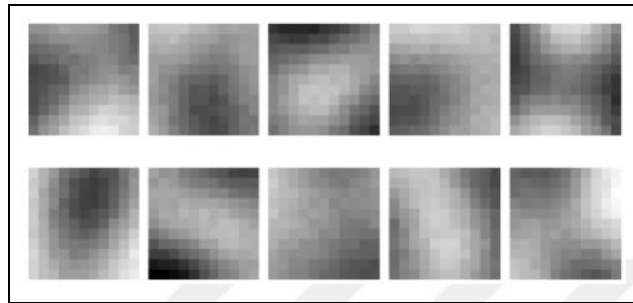
Figure 4.6. The convolutional filters of the preprocessed cross-sectional volumes

The best prognostic power was achieved by the extracted features of the preprocessed longitudinal volumes with a patch size of 15, an autoencoder with 12 hidden units and a significance level of 0.005. As a result, the learnt 12 filters are shown in Figure 4.7 below.
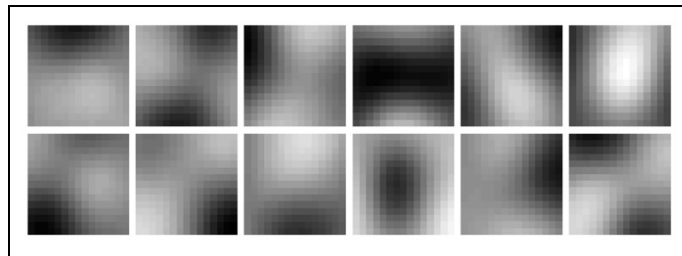
Figure 4.7. The convolutional filters of the preprocessed longitudinal volumes

4.2.3. Feature Maps

In this section, we provide two feature maps, each of which is obtained by convolving two preprocessed volumes of different groups with the same convolutional filter. Figure 4.8 (a) presents the an axial slice of the preprocessed cross-sectional volume of an MCI patient who developed AD (a), an MCI patient who remained stable (c), and the acquired feature maps (b, d) for these slices, respectively, in order to show the effect of same filter on different images.
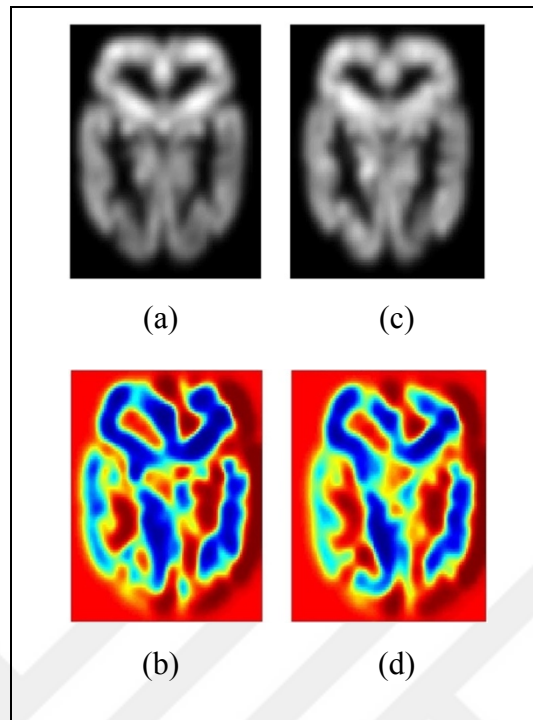
Figure 4.8. Preprocessed cross-sectional volumes and their corresponding feature maps. (a) Belongs to convert-MCI group, (b) and its feature map, (c) Belongs to nonconvert-MCI group, (d) its feature map

4.3. FEATURE SELECTION

To tackle with the high-dimensionality of the acquired features, we employed feature selection procedure before performing classification. We selected different feature selection methods for the features extracted from quantitative data and volume data due to their nature. For the quantitative data, we selected a feature selection algorithm based on statistical tests Additionally, a SVM based algorithm is employed for the volume data.

4.3.1. Preprocessed Quantitative Data

The high-dimensionality of neuropsychological assessment data is reduced by a classifier-independent feature subset selection method based on a non-parametric statistical hypothesis test, where the importance of a feature is determined by a p-value from the Wilcoxon rank-sum test statistics as given in Table 4.8. An asterisk (*) indicates a significance level of 0.05 and two asterisks (**) are used to denote the significance level as 0.01.

Table 4.8. The Wilcoxon rank sum statistics for all pairs of diagnostic groups

	ARCD MCI	ARCD AD	ARCD VD	MCI AD	MCI VD	AD VD
Orientation						
Personel information	0.24	*	0.33	**	0.98	0.07
Orientation	**	*	**	0.18	0.94	0.19
Attention-executive functions						
Digit span forward	0.13	**	**	0.16	0.39	0.68
Digit span backward	*	*	**	0.06	0.55	0.28
Mental control	0.53	**	**	0.07	**	0.44
Semantic fluency	*	*	*	0.07	0.16	0.62
Fonemic fluency	**	*	*	0.27	**	0.37
Similarities	**	*	**	*	0.71	0.06
Proverbs	0.12	*	*	0.07	0.10	0.86
Clock drawing	0.74	*	**	*	**	0.88
Stroop interference	*	*	*	0.88	**	**
Stroop incorrect responses	*	0.07	*	0.47	0.66	0.36
Stroop spontaneous	0.29	0.08	**	0.32	**	0.24
Language						
Naming (BNT)	**	*	*	**	0.34	0.53
Semantic fluency	*	*	*	0.07	0.16	0.62
Fonemic fluency	**	*	*	0.27	**	0.37
Visuo-spatial functions						
Face recognition	0.49	*	0.08	0.14	0.43	0.60
Copying drawings	0.10	0.63	0.18	**	*	0.44
Memory functions						
WMS-R	*VR immediate recall*	*	*	*	**	0.12	0.49
	VR delayed recall	*	*	*	**	0.35	*
	VR recognition	**	*	**	**	0.77	*
	LM immediate recall	*	*	0.46	**	0.07	*
	LM delayed recall	*	*	0.49	0.06	*	*
Ö-VMP	*Immediate memory*	**	*	*	0.12	0.27	0.51
	Total learning score	*	*	*	**	0.79	**
	Highest learning score	*	*	*	*	0.36	*
	Delayed free recall	*	*	*	*	**	*
	Recognition score	*	*	*	*	0.13	*

The ARCD and AD pair as it is shown in Table 4.8, was composed of all neurocognitive tests except WMS-R digit span forward (p<.05), WMS-R mental control (p<.05), Stroop test incorrect responses, Stroop test spontaneous corrections and copying drawings tests. The ARCD and MCI pair didn't consider the test scores measures orientation and visuo-spatial abilities in the training of machine learning algorithms. Similarly, the ARCD and VD pair excluded the test scores of orientation and visuo-spatial categories. Face recognition, similarities and clock drawing only considered by the ARCD and AD pair.

On the other hand, the most statistically significant overlap occurred between the MCI-VD pair. All neurocognitive test scores of orientation and language functions are statistically insignificant. Only copying drawings and WMS-R logical memory delayed recall test scores have significance in the MCI-VD pair. Impaired face recognition is one of the neurocognitive hallmarks for differentiating cognitive impairment for AD patients [148]. Face recognition test scores were statistically significant only in the ARCD-AD pair. WMS-R personal-current information and WMS-R orientation features were significantly different only in the case of the differential diagnosis of ARCD from AD.

4.3.2. Preprocessed Volume Data

Since we had 65 number of clusters for the longitudinal data, we came up with a high-dimensional feature vector depend on the number of hidden units in the trained autoencoder. For example, when we trained an autoencoder with 12 hidden units using the longitudinal data, we acquired a 780 dimensional feature vector. There was no need to apply any feature selection method due to few clusters obtained for the cross-sectional case.

For this purpose, we utilized the support vector machine based recursive feature elimination method (SVM-RFE) [144] implementation of Weka software. We sorted the acquired features based on their calculated SVM-RFE scores. In each iteration, we eliminated the feature having the least SVM-RFE score, and then, we evaluated their classification performances for the remaining ones. The accuracy of the total system increased and peaked at some point as we eliminated the least important one. We considered the remaining features at the peak point as our selected subset of the features to assess the predictive power of them using machine learning algorithms.

4.4. CLASSIFICATION PERFORMANCES

The classification performances were assessed differently for three kind of datasets of this thesis to test our hypothesis.

4.4.1. Predictive Power of Neuropsychological Test Scores

A multilayer perceptrons, a support vector machine with a linear kernel and a M5-model tree classifiers were employed to assess the predictive powers of the neuropsychological test scores for distinguishing age-related cognitive decline from other dementias.

The following Table 4.9 shows the performances of three classifiers trained using all neuropsychological test scores of the patients in terms of accuracy (ACC), sensitivity (TPR), specificity (SPC), precision (PPV), F-measure (F1) and area under the ROC curve (AROC).

Table 4.9. Classification results for distinguishing ARCD based on all test scores

	ACC	*TPR*	*SPC*	*PPV*	*F1*
ARCD vs MCI					
MLP	73.8	80	68.6	74.7	73.8
SVM	76.9	73.3	80	76.9	76.9
M5P	*90.8*	*96.7*	*85.7*	*91.5*	*90.8*
ARCD vs AD					
MLP	84	75	90	84	83.8
SVM	88	80	93.3	88.1	87.9
M5P	*100*	*100*	*100*	*100*	*100*
ARCD vs VD					
MLP	74.5	83.3	61.9	74.3	74.1
SVM	72.5	80	61.9	72.3	72.3
M5P	*80.4*	*86.7*	*71.4*	*80.3*	*80.2*

Similarly, the same classifiers were trained using only selected subsets of the neuropsychological test scores and the classifier performances are shown in Table 4.10 in terms of same measurement types. The best performances among three different type of algorithms are marked in bold. Diagnostic predictability was over 80% for all treatment

groups. When Table 4.9 and Table 4.10 are compared, it is observed that eliminating insignificant neurocognitive test scores resulted in a slightly increased performance in terms of accuracy.

Table 4.10. Classification results of the pairs of ARCD using selected features

	ACC	*TPR*	*SPC*	*PPV*	*F1*
ARCD vs MCI					
MLP	89.2	90	88.6	89.3	89.2
SVM	89.2	86.7	91.4	89.2	89.2
M5P	*95.4*	*96.7*	*94.3*	*95.4*	*95.4*
ARCD vs AD					
MLP	86	80	90	85.9	85.9
SVM	88	80	93.3	88.1	87.9
M5P	*100*	*100*	*100*	*100*	*100*
ARCD vs VD					
MLP	78.4	83.3	71.4	78.3	78.3
SVM	*86.3*	*93.3*	*76.2*	*86.5*	*86.1*
M5P	84.3	90	76.2	84.3	84.2

Table 4.11 shows the neurocognitive test scores with the higher merit values among the subset of selected neurocognitive test scores. The worthiest feature for the ARCD-AD pair was "Öktem-VMPT recognition score". Additionally, Öktem-VMPT total learning score has the higher merit value for M5P classifier that has performed better among all three classifiers in the ARCD-MCI pair.

Another observation is that the neuropsychological test scores with a higher merit value in the best classification model of the age-related cognitive decline and vascular dementia pair were semantic fluency, proverbs, Stroop interference and Boston naming test.

In order to compare the performances of machine learning algorithms we utilized statistical tests as we mention in Chapter 3. The Cochran's Q test and McNemar's test statistics are presented in Table 4.12, where the obtained values greater than the critical value ($\chi^2_{(1,0.95)}$=3.841) for the McNemar's test and the obtained Q values greater than the critical value ($Q_{(2,0.95)}$= 5.99) for the Cochran's Q tests are indicated in bold.

Table 4.11. The merit features of the classifiers for each of diagnostic pairs

MLP	SVM	M5P
ARCD vs MCI		
Semantic fluency	WMS-R digit span backward	Ö-VMPT total learning
Stroop interference	Semantic fluency	
WMS-R LM immediate recall	Stroop interference	
Ö-VMPT highest learning	Stroop incorrect responses	
Ö-VMPT delayed free recall	WMS-R LM delayed recall	
Ö-VMPT recognition	Ö-VMPT total learning	
	Ö-VMPT delayed free recall	
	Ö-VMPT recognition	
ARCD vs AD		
Ö-VMPT recognition	Ö-VMPT recognition	Ö-VMPT recognition
ARCD vs VD		
Semantic fluency	Semantic fluency	Proverbs
Fonemic fluency	Proverbs	Stroop interference
Proverbs	Stroop interference	Ö-VMPT total learning
Stroop interference	Naming BNT	Ö-VMPT delayed free recall
WMS-R VR immediate recall		
Ö-VMPT immediate memory		
Ö-VMPT delayed free recall		

Table 4.12. McNemar's and Cochran's Q test statistics to compare classifiers

	MLP vs SVM		MLP vs M5P		SVM vs M5P		All	
	χ	p	χ	p	χ	p	Q	p
ARCD vs MCI	0.25	0.62	1.13	0.29	1.13	0.29	3.2	0.2
ARCD vs AD	0	1	**5.14**	0.02	**4.17**	0.04	**9.55**	0.01
ARCD vs VD	2.25	0.13	0.57	0.45	0	1	2.89	0.24

The results demonstrate that all three classifiers can be successfully used for distinguishing age-related cognitive decline from other dementias. Beside this, the M5-model regression tree classifier performs better for distinguishing age-related cognitive decline from Alzheimer's diseases in all cases.

4.4.2. Predictive Power of Cross-sectional Data

Table 4.13 presents the classification performances of the prognostic cross-sectional features using the following classifiers: SVM with polynomial kernel (SP), SVM with linear kernel (SL) and logistic regression (LR) in terms of accuracy, sensitivity and specificity. The diagnostic predictability resulted in an accuracy between 73.4%-78.7% for all classification models using ten-fold cross-validation. The best performance was achieved with a configuration of 10 feature detectors (#f) and five volume-of-interests (#v) at the significance level of 0.005.

Table 4.13. The classification performances of the acquired features

#f	#v	Accuracy (%)			Sensitivity (%)			Specificity (%)		
		SP	SL	LR	SP	SL	LR	SP	SL	LR
10	5	78.7	73.1	76.1	80.6	77.6	78.8	76.4	67.9	72.9
10	7	75.7	70.2	74.1	80.6	78.2	72.1	70.0	60.7	75.8
25	5	76.7	73.8	74.8	78.8	75.8	75.8	74.3	71.4	73.6
25	7	73.8	73.8	65.6	78.2	76.4	66.1	68.6	70.7	65.0
50	5	76.1	77.4	74.4	77.6	78.8	76.4	74.3	75.7	72.1
50	7	73.4	73.4	74.1	75.8	77.0	77.0	70.7	69.3	70.7

In order to have a better evaluation of the proposed method, we compared our results with two other conventional methods. The first comparative method was that taking the all significant voxels of the five clusters as features to train classifiers, which is mentioned as "without VBM" in the Table 4.14. Another comparative method employs an additional dimensional reduction step to the "without VBM" method based on principal component analysis (PCA), which is called as "without VBM+PCA" in the table.

A support vector machine with polynomial kernel, a support vector machine with linear kernel and a logistic regression were used as classifier algorithms. The results were compared with our best result labelled as "novel CNN" in terms of accuracy, sensitivity and specificity. To sum up, the comparison showed that the novel convolutional neural network model introduced in this thesis extracted features of cross-sectional data with a high predictive power than other two comparative methods.

Table 4.14. Performance comparison of the proposed solution.

	Accuracy (%)			Sensitivity (%)			Specificity (%)		
	SP	SL	LR	SP	SL	LR	SP	SL	LR
Novel CNN	78.7	73.1	76.1	80.6	77.6	78.8	76.4	67.9	72.9
Without *VBM*	63.3	64.3	65.6	67.3	68.5	70.3	58.6	59.3	60
Without *VBM+PCA*	63.9	66.9	74.1	64.2	67.3	77	63.6	66.4	70.7

4.4.3. Predictive Power of Longitudinal Data

The prediction performances of all and selected subset of features extracted from the preprocessed longitudinal volumes are presented in Table 4.15 in terms of accuracy (ACC) sensitivity (TPR), specificity (SPC), precision (PPV) and F-measure (F1). Three classifiers were employed to assess the predictive power of the features using a ten-fold cross validation procedure, which those classifiers are a support vector machine (SVM), a linear discriminant analysis (LDA) and a multilayer perceptron (MLP) using with configurations of patch size and number of filters.

The performances above 90% in all statistical measures are marked with *bold* and have a grey background. Other statistical measures above 90% are marked only with *bold*. Furthermore, the highest performance is marked with blue that was achieved when we trained an autoencoder with 12 hidden unit using patches in size of 15. This combination yielded an accuracy of 97% with a sensitivity of 96.8% and a specificity of 97.3% with a subset of features determined by SVM-RFE method.

The best performance was achieved in the configuration using patches of size 15×15, 12 filters, a subset of 99 prognostic features and a support vector machine classifier, where an accuracy of 97%, a sensitivity of 96.8% and a specificity of 97.3% was achieved.

It is worth to mention that the support vector machine classifier always produced the best results among the three classifiers. Furthermore, the patches in size of 8 couldn't extract enough information from the longitudinal data to differentiate the diagnostic groups. The better performances were achieved when the patches are sampled in size of 13 and 15 with an accuracy above 90%.

Table 4.15. The results of the CAD system for different configurations

Patch size Filters	Classifier	Total set of prognostic features					Subset of prognostic features				
		ACC	TPR	SPC	PPV	F1	ACC	TPR	SPC	PPV	F1
15×15 14	LDA	67.3	65.8	68.4	67.6	67.4	85.1	87.4	82.2	85.1	85.1
	SVM	78	69.9	84.2	77.9	77.8	**94.6**	**96.8**	**91.8**	**94.7**	**94.6**
	MLP	76.8	71.2	81.1	76.7	76.7	89.3	89.5	89	89.3	89.3
15×15 12	LDA	75	69.9	78.9	74.9	75	88.1	90.5	84.9	88.1	88.1
	SVM	81.5	76.7	85.3	81.5	81.5	**97**	**96.8**	**97.3**	**97**	**97**
	MLP	78	74	81.1	77.9	78	**95.2**	**95.8**	**94.5**	**95.2**	**95.2**
15×15 10	LDA	72.6	71.2	73.7	72.8	72.7	79.8	82.1	76.7	79.8	79.8
	SVM	81.5	76.7	85.3	81.5	81.5	**91.7**	**95.8**	86.3	**91.8**	**91.6**
	MLP	75	69.9	78.9	74.9	75	86.9	86.3	87.7	87.1	86.9
15×15 8	LDA	61.9	60.3	63.2	62.3	62	67.9	72.6	61.6	67.8	67.8
	SVM	75	74	75.8	75.2	75.1	84.5	83.2	86.3	84.8	84.6
	MLP	69.6	64.4	73.7	69.6	69.6	79.2	78.9	79.5	79.4	79.2
13×13 14	LDA	73.8	69.9	76.8	73.8	73.8	86.3	89.5	82.2	86.3	86.3
	SVM	78	74	81.1	77.9	78	**92.3**	**93.7**	**90.4**	**92.3**	**92.3**
	MLP	75	71.2	77.9	75	75	88.7	**91.6**	84.9	88.7	88.7
13×13 12	LDA	78.6	71.2	84.2	78.5	78.4	87.5	91.6	82.2	87.5	87.5
	SVM	79.8	72.6	85.3	79.7	79.6	**93.5**	**94.7**	**91.8**	**93.4**	**93.4**
	MLP	76.2	69.9	81.1	76.1	76.1	88.1	91.6	83.6	88.1	88.1
13×13 10	LDA	73.2	74	72.6	73.7	73.3	85.1	86.3	83.6	85.1	85.1
	SVM	76.2	71.2	80	76.1	76.1	**92.3**	**95.8**	87.7	**92.4**	**92.2**
	MLP	77.4	71.2	82.1	77.3	77.3	89.3	**91.6**	86.3	89.3	89.3
13×13 8	LDA	64.9	67.1	63.2	65.7	65	77.4	81.1	72.6	77.3	77.3
	SVM	73.8	63	82.1	73.7	73.5	83.9	**90.5**	75.3	84.1	83.8
	MLP	73.8	68.5	77.9	73.7	73.8	80.4	85.3	74	80.3	80.3
11×11 14	LDA	76.2	76.7	75.8	76.5	76.3	76.8	81.1	71.2	76.7	76.7
	SVM	81.5	78.1	84.2	81.5	81.5	86.3	89.5	82.2	86.3	86.3
	MLP	71.4	60.3	80	71.4	71.1	85.7	89.5	80.8	85.7	85.7
11×11 12	LDA	75.6	71.2	78.9	75.6	75.6	88.7	89.5	87.7	88.7	88.7
	SVM	83.9	79.5	87.4	83.9	83.9	**95.8**	**95.8**	**95.9**	**95.8**	**95.8**
	MLP	78.6	72.6	83.2	78.5	78.5	**92.3**	**93.7**	**90.4**	**92.3**	**92.3**
11×11 10	LDA	75	71.2	77.9	75	75	87.5	89.5	84.9	87.5	87.5
	SVM	79.2	75.3	82.1	79.1	79.1	**92.3**	**93.7**	**90.4**	**92.3**	**92.3**
	MLP	76.8	69.9	82.1	76.7	76.7	88.7	**91.6**	84.9	88.7	88.7
11×11 8	LDA	72.6	69.9	74.7	72.7	72.7	80.4	81.1	79.5	80.5	80.4
	SVM	72	63	78.9	71.9	71.8	84.5	87.4	80.8	84.5	84.5
	MLP	70.2	58.9	78.9	70	69.9	84.5	89.5	78.1	84.6	84.4

Figure 4.9 presents the accuracies obtained each feature elimination steps for the best case achieved with a patch size of 15 and a filter number of 12. During the feature elimination process, the accuracy is higher when only 99 features are selected.
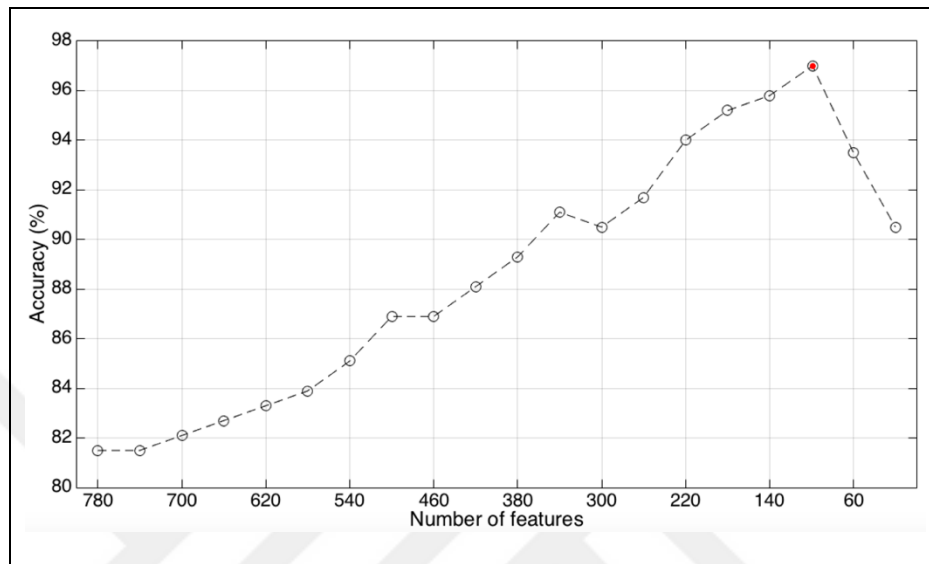
Figure 4.9. The accuracies of the feature elimination steps

In order to measure the performance of the preprocessed longitudinal volume, we compared the results of the CAD system for the best configuration in terms of number of filters and patch size for the baseline only and the 12-months follow up exam only. For every case a different SPM-F contrast map is generated.

Table 4.16. Comparison of the baseline, the follow-up and the preprocessed longitudinal

Patch size Filters	Baseline			Follow-up			Preprocessed Longitudinal		
	ACC	TPR	SPC	ACC	TPR	SPC	ACC	TPR	SPC
15×15 12	**91.1**	90.4	91.6	**93.5**	90.8	95.8	**97**	96.8	97.3

As shown in Table 4.16, the preprocessed longitudinal volume performed better than the only baseline or the only follow-up thus proving that the differences between the two exams contain significant information for the prediction prognosis. The configuration with the

baseline exam only performed an accuracy of 91.1% and with the 12-months follow-up only resulted in an accuracy of 93.5%.

4.5. SUMMARY OF THE CHAPTER

In this chapter, we presented the results on the experimental evaluation of the hypothesis in four main sections in correspondence with the four main stages of the methodology explained in Section 3.6. Firstly, the descriptive statistics of the datasets were provided. Then, information on extracted features are given. Next, experimental set-ups are provided. Finally, the predictive capacities of the acquired features are provided.

In order to test our hypothesis, we obtained three different datasets. A neuropsychological assessment data set consisting of 26 neuropsychological test scores of 106 patients with age related cognitive decline, Alzheimer's disease, vascular dementia or mild-cognitive impairment was obtained from the neurology department of Maltepe University Hospital. Other dataset was composed of a set of structural MRI data that scanned in the baseline examination of two groups of patients with MCI consisting of a group of patients who remained stable in MCI and a group of patients who subsequently developed AD. Additionally, we obtained another dataset of the longitudinal structural MRI data of MCI patients who eventually developed AD or not. The patient groups were statistically significantly similar on age, education and gender (p<0.001) in each dataset.

Then, we presented the feature extraction methods applied to convert our datasets into feature vectors. In this thesis, we used the neuropsychological test scores of a patient directly as the feature vector for that patient. However, we introduced a novel single-layer convolutional neural network schema to represent the MRI data of a patient as a feature vector.

Next, we employed feature selection methods to tackle with the high-dimensionality of the feature vectors acquired from the neuropsychological assessment data. A classifier-independent feature subset selection method based on a non-parametric statistical hypothesis test were used, where the importance of a feature is determined by the p-value from the Wilcoxon rank-sum test statistics. Selecting a classifier-independent feature selection

method provided a fair comparison between the classification algorithms. Evaluating the statistical significance of the selected features leads to a model that is more understandable according to the statistical methods used by physicians in the neuropsychological domain. In this chapter, we presented selected neuropsychological tests based on Wilcoxon rank-sum test statistics.

Finally, in this chapter, we presented the experimental results on evaluating the predictive powers of the acquired feature vectors on each datasets. Those experimental setups aimed to test our hypotheses of this thesis.

We designed six experimental setups for each pairwise classification models with the combination of three different classification algorithms (multilayer perceptrons, support vector machines and M5-model trees) and two sets of features (all 26 test scores and selected test scores) for the CAD of age-related cognitive decline based on neuropsychological assessment data. The results of the experiments based on all 26 test scores allowed us to accept the hypothesis-I that says "age-related cognitive decline can be distinguished from other dementias using machine learning algorithms based on neuropsychological assessment data". To be more specific, age-related cognitive decline can be distinguishable from Alzheimer's disease with an accuracy of 100% (a sensitivity of 100%; a specificity of 100%), from mild-cognitive impairment with an accuracy of 90.8% (a sensitivity of 96.7%; a specificity of 85.7%) and from vascular dementia with an accuracy of 80.4% (a sensitivity of 86.7%; a specificity of 71.4%) as it is seen in Table 4.9. Additionally, the results of the experiments based on only selected test scores (as in Table 4.10) allowed us to accept the hypothesis-II that says "if a feature selection method is used to select a subset of neuropsychological test scores, the performances of the classifiers aimed to identify age-related cognitive decline will improve". The improved results were reported in this thesis as an accuracy of 95.4% (a sensitivity of 96.7%; a specificity of 94.3%) for MCI and an accuracy of 86.3% (a sensitivity of 93.3%; a specificity of 76.2%) for VD.

Furthermore, the six configurations were designed with the combination of one patch size (11x11), three different number of filters (10, 25, 50) and two different significance levels (0.001% and 0.005%) for the CAD of MCI-due-to-AD based on baseline MRI. Among those, we reported an accuracy of 78.7% (a sensitivity of 80.6%; a specificity of 76.4%) to

predict the risk of having AD for MCI patients based on their structural baseline MR scans, which is 6% more than the reference study [70]. Furthermore, we conducted another set of experiments to test the effect of our proposed pooling method. So, we have achieved more than 10% accuracy compared to the comparative methods, which lead us to accept the hypothesis-V that claims the importance of considering anatomical structure of brain on pooling.

Similarly for the preprocessed longitudinal volumes, the combination of three different patch sizes (11x11, 13x13, 15x15) and four different number of filters (8, 10, 12, 14) resulted in 12 experiments for the evaluation of the prognostic power of the preprocessed longitudinal volumes. Thus, we accepted the hypothesis-IV based on the reported results in Table 4.15 with an accuracy of 97% (a sensitivity of 96.8% and a specificity of 97.3).

5. DISCUSSION

In this thesis, we aimed to distinguish age-related cognitive decline from other dementias and to predict the risk of having Alzheimer's disease for MCI patients. The following subsections discuss the finding of this thesis in that two perspectives.

5.1. DISTINGUISHING AGE-RELATED COGNITIVE DECLINE

In this thesis we attempted to differentiate age-related cognitive decline from other dementias by utilizing three different commonly used machine learning algorithms (multilayer perceptrons, support vector machines and M5-model trees). As it is explained in Chapter 3, it is essential to predict the abnormal decline in cognitive abilities that is not related to normal aging in an early stage. Concerning the neurocognitive tests, traditionally, a neurocognitive assessment is performed as a set of tests aiming to examine several cognitive ability areas such as orientation, attention, executive functions, visuo-spatial skills, language and memory [149]. The difficulties in the clinical interpretation of neurocognitive data for differential diagnosis of dementia have raised the use of computational techniques.

The neurocognitive test battery used in this thesis was composed of several tests that covers all of these cognitive domains. We trained our pairwise classification models based on those huge dataset. In this regard, the classification models were totally distinguished age-related cognitive decline from Alzheimer's disease with an accuracy of 100%.

In studies using machine learning approaches, adequate feature selection is an important task for the creation of a classification model that will have a successful interpretation of data, a reduced variance and an improved classification accuracy in general [150]. For that reason, we have used a classifier-independent feature subset selection method based on a non-parametric statistical hypothesis test, where the importance of a feature is determined by a p-value from the Wilcoxon rank-sum test statistics of the difference between groups of the classifier model. Selecting a classifier-independent feature selection method provided a fair comparison between the classification algorithms and improved our results up to an accuracy of 95.4% (a sensitivity of 96.7%; a specificity of 94.3%) for ARCD vs. MCI pair.

Furthermore, evaluating the statistical significance of the selected features leads to a model that is more understandable according to the statistical methods used by physicians in that domain.

All methods presented a high success rate in their predictions with the M5P classifier performing better for differentiating ARCD from other dementias. Additionally, other popular machine learning algorithms, like SVM and MLP classifiers, achieved high classification performances.

We can conclude that the neurocognitive tests are sufficient enough to differentiate ARCD from MCI or AD or VD. Additionally, neurocognitive tests of different diagnostic groups were pairwise compared accompanied with a statistical analysis to individual tests in order to measure their statistical importance. According to the findings it is possible to use computational approaches in neurocognitive assessment in clinic after eliminating insignificant test scores.

Additionally, a decision-making system of diagnosis can be enriched with data originating from different sources such as electroencephalogram and neurocognitive tests [151]. For that reason, even if the analysis of our neurocognitive test data shows some significant results, additional data obtained from other sources such as PET, MRI, fMRI or biomarkers could furthermore contribute to the differential diagnosis.

5.2.   PREDICTING THE RISK OF AD FOR MCI PATIENTS

Until today a treatment that can cure Alzheimer's disease or any other type of dementia is still to be discovered. Thanks to recent advances in medical technology, drug and non-drug treatments are discovered and validated in patients with Alzheimer's disease in routine clinical use, thereby improving their quality of life and preventing the progression of symptoms in the early stages of dementia. On the other hand, the majority of those patients are older patients. Therefore, drug treatment in older patients should be carefully considered in terms of adverse effects with eventual other medications of the patients. Thus, the possibility of identifying individuals with MCI that will progress to AD would be very beneficial for the patients in terms of good prognosis and delaying/preventing the symptoms

of Alzheimer's disease by immediately applying a neuro-modulation treatment, and a treatment that slows down the progression to AD. This thesis aimed to help patients by reducing the use of unnecessary medications for individuals who are predicted to remain stable in MCI.

Table 5.1. Recent studies in the literature aiming to predict the risk of developing AD

Studies	*Longitudinal*	*Invasive*	*ADNI*	*Data type/* *Number of Patients*	*ACC*	*TPR*	*SPC*
Current study	**Yes**	**No**	**Yes**	**MRI** **73 C-MCI, 95 NC-MCI**	**97%**	**96.8%**	**97.2**
Devanand *et al* [84]	No	No	No	MRI+Clinical+Cognition 27 C-MCI, 79 NC -MCI	94.9%	92.6%	85.2%
Green *et al* [117]	No	Yes	No	EEG+CSF 19 C-MCI, 17 NC-MCI	94.4%	94.7%	94.1%
Nazeri *et al.* [118]	Yes	Min	Yes	MRI+Plasma 110 C-MCI, 190 NC -MCI	93.5%	N/A	N/A
Douaut *et al.* [119]	No	Yes	No	MRI(DTI)+CSF 13 C-MCI, 22 NC -MCI	91%	85%	96%
Minhas *et al.* [83]	Yes	No	Yes	MRI+Cognition N/A	89.7%	87.5%	92.3%
Peters *et al.* [87]	No	No	No	MRI+Cognition 18 C-MCI, 22 NC -MCI	87.5%	90.9%	83.3%
Dukart *et al.* [120]	No	Min	Yes	MRI+PET+Genetic 177 C-MCI, 265 NC -MCI	86.8%	87.5%	77.8%
Ortiz *et al.*[121]	No	No	Yes	MRI N/A	83%	67%	95%
Ardekani *et al.* [86]	Yes	Min	Yes	MRI+Cognition+Genetic 86 C-MCI, 78 NC -MCI	82.3%	86%	78.2%
Korolev *et al.* [122]	No	Yes	Yes	MRI+Clinical+CSF 139 C-MCI, 120 NC -MCI	80%	83%	76%
Willette *et al.* [85]	No	No	Yes	MRI+Cognition 76 C-MCI, 86 NC -MCI	80%	78.3%	81.5%
Suk *et al.* [70]	No	No	Yes	MRI+PET 76 C-MCI, 128 NC-MCI	75.9%	48%	95.2%
Suk *et al.* [70]	No	No	Yes	MRI 76 C-MCI, 128 NC-MCI	72.4%	36.7%	91%

It is generally accepted that using non-invasive data is preferable for dementia patients because the use of invasive methods may arise ethical issues for patients with impaired

mental function. The techniques for obtaining CSF measurements are full invasive methods that may be painful or may have post-operation complications. In addition, those methods should be performed under anaesthesia. General anaesthesia carries a higher risk for the elderly people and local anaesthesia may not be applicable for them due to several health conditions. Furthermore, PET or MRI scans with a contrast agent can be considered as a minimally invasive method due to administration of a contrast agent. Thus, we have concentrated on developing CAD systems based on non-invasive data. In the studies presented in Chapter 4 and 5, we utilized T1-weighted MR images that are totally non-invasive data.

In this thesis, we introduced two CAD systems for MCI patients based on their cross-sectional and longitudinal structural MRI data to predict their risk for developing AD eventually. We compared our models with similar systems that present one or more common characteristic such as the use of longitudinal data or not, data from ADNI or another database, and the use of invasive minimal invasive or non-invasive methods. Thus, Table 5.1 presents the most successful recent studies in terms of accuracy, sensitivity and specificity. Devanand *et al.* [84] proposed a non-ADNI study with high accuracy results (94.9%) but with relatively low specificity (85.2%) combining MRI, clinical and cognition data obtained in the baseline. Green *et al.* [117] proposed a CAD system with a very high accuracy of 94.4% (94.7% sensitivity, 94.1% specificity) based on electroencephalography (EEG) test scores and CSF data (an invasive method) of a non-ADNI dataset. In both studies, the number of C-MCI patients was relatively small.

Among the studies that utilized the ADNI dataset, the study of Nazeri *et al.* achieved an accuracy of 93.5% using longitudinal data of the MRI and plasma protein data [118] which is a minimal invasive method. The study of Douaut *et al.* was interesting due to the employment of diffusion tensor imaging (DTI) data, which are not available in ADNI dataset [119]. When comparing our study in terms of ADNI dataset and the use of invasive methods or additional tests, the closest among the studies of the Table 5.1 is the one of Minhas *et al.* [83] which has lower success rates (accuracy 89.7%) and still uses cognitive tests.

We compared our results based on the cross-sectional data with the study of Suk *et al.* [70] as a reference study. Unlike the study of Suk *et al.* [70], which used two-layered deep

network, we only used a single-layer convolutional neural network. Our approach resulted in a slightly superior accuracy of 78.7% than that of the study of Suk *et al.* (more than 6%). Additionally, we compared the success of the proposed method with two additional methods: one taking in account the all significant grey matter voxels and another using principal component analysis for reducing them. The proposed method performed better than the other comparative methods more than 10% in accuracy.

As a general observation, the CAD model of longitudinal data presents slightly higher results when compared with similar systems that use invasive methods or clinical and cognition tests in order to achieve a high performance. Moreover, it performed much better when compared with systems where no additional tests or invasive methods were utilized. As we said before in the introduction, the use of non-invasive method for acquiring data in elderly people and patients with dementias is very essential. Our approach achieved high success rates with an accuracy of 97%, a sensitivity of 96.8% and a specificity of 97.3%. When comparing the results of the proposed method with the studies listed in Table 5.1 we see that this system outperforms them all.

In neuroimaging, the enhanced performance of longitudinal data over cross sectional data has been validated by numerous studies: Xu *et al.* [152] demonstrated that for the detection of brain imaging phenotypes, longitudinal data outperform cross-sectional data. For neurodegenerative diseases, longitudinal ADNI studies [83], [86], [118] performed better than the cross-sectional studies [85], [106], [122]. According to the above observations, the exploration of longitudinal data and the research for new methods in order to extract information from them could be beneficial for the prognosis and treatment of dementias.

In comparison with other systems, the acquired results were higher from all known systems and were comparable only with those that utilized invasive methods like CSF tests [117], [119], minimal invasive methods like plasma [118] or additional cognitive tests [84]. In particular, the CSF test is one of the diagnostic methods for brain or spine related diseases. During this test, a sample of cerebrospinal fluid is removed in a procedure called "Lumbar puncture" often performed in anaesthesia. The CSF sampling is performed by inserting a needle into the spinal canal where several trials may be needed with different sizes of needles [153] thus elevating the risk of back pain and trauma. In addition, lumbar puncture may have

some post-operation complications like cerebral and spinal herniation post lumbar puncture headache [154].

From that point of view, the proposed method presents a significant advantage as it does not need invasive procedures which may be uncomfortable, inappropriate or tests that sometimes are difficult to be executed by the elderly people [155].

Using a system for the prognostic diagnosis of Alzheimer's disease in MCI patients is vital to start the treatment immediately using preventive or therapeutic medicines for slowing down the progression to AD or treating the symptoms of dementia [156]. However, the benefit of medication treatment for preventive purposes should be considered carefully in a geriatric population due to the comorbidities and side-effects of drugs. The use of preventive medication in MCI patients who actually will not develop AD may expose those patients to the unnecessary burden of medication. For this reason, the identification of MCI patients who are most likely to convert to AD with a high accuracy is essential in terms of preventing unnecessary medication-related problems.

The CAD system yielded a 97% accuracy (96.8% sensitivity and 97.2% specificity), thus contributing to the efforts towards the creation of a prognosis prediction system which will decide about the type of treatment of mild cognitive impairment patients without necessarily involving the use of invasive methods or cognitive tests.

Furthermore, this thesis has two main methodological contributions to the literature. The one of them is about the pooling layer of the convolutional neural network. Traditional pooling methods define the grid-shape pooling regions. In this thesis, we proposed a novel pooling method. The idea behind this proposed pooling method is the spatial relevance of brain tissues in neighbouring voxels. For that purpose, we used the significant clusters (volume-of-interests) defined by the voxel-based morphometric analysis results as the regions of pooling. In the study presented in Chapter 4, this method is applied successfully using the five clusters at significance level of 0.005 and the seven clusters at significance level of 0.001. In the study presented in Chapter 5, we have improved our idea by associating the clusters with individual regions of the brain, thus, we achieved better results in terms of accuracy. In other words, an atlas-based region-of-interest (ROI) analysis was performed on

the SPM-F contrast map using the "TD Labels" atlas of the Wake Forest University (WFU) PickAtlas to determine which voxels will be pooled together.

The other methodological contribution of this thesis focus on the representing longitudinal data in a single-volume. We have proposed a novel method to merge baseline and 12-months follow-up images as it is presented in Chapter 5. The method gets benefit of progressive grey matter atrophy by modelling the difference between two time-point images using a method based on the jacobian determinant. To acquire the representative longitudinal volume, we got the slice-wise jacobian determinant of the baseline image and follow-up image and added the result to the follow-up image for each patient. Comparison of the classification performances between baseline only or follow-up only exams showed that the follow-up exams have more discriminative information than the baseline. This is an expected situation in biological terms, because the progressive MCI patients have higher atrophy rate than the stable MCI patients.

5.3. SUMMARY OF THE CHAPTER

In this thesis, we explored the applicability of deep learning approaches to extract features of neuroimaging data for predicting the risk of developing Alzheimer's disease for mild cognitive impairment patients. Beside this purpose, we also explored the distinguishability of age-related cognitive decline from Alzheimer's disease, vascular dementia and mild cognitive impairment based on only neuropsychological test scores. The traditional classification models of neuroimaging generally depend on neuroimaging on hand-crafted features. Some hand-crafted features requires excessive domain knowledge and expertise or may include bias [46]. Thus, we considered the use of convolutional neural networks that are successfully used in the field of image processing [157]. We especially dealt with the possibility of extracting features based on unsupervised training. For this purpose, we employed a patch-based autoencoder to learn a set of feature detectors to be used in the convolutional layer. A large amount of two-dimensional patches are randomly selected from the structural MRI images of the patients with a baseline diagnosis of MCI. Then, those patches are vectorised to create a training dataset for and autoencoder. After the training procedure of the autoencoder, we have obtained a set of feature detectors that are sensitive to the informative parts of the structural MRI images. The acquired feature detectors were

employed in the convolutional neural network model as convolutional kernels to extract the neuroimaging features of the patients.

Furthermore, we had two methodological contribution in this thesis. The first one is a new pooling approach for the subsampling layer of CNN based on VBM analysis that selects regions with a statistical significance between converted and non-converted patients with MCI. When comparing our results with similar studies that did not use our pooling approach, we observed that our method performed better (78.7%), thus showing that the VBM analysis can be integrated successfully with deep learning approaches for MCI prediction prognosis. The second one is a novel information extraction method of longitudinal MRI data aiming to "emphasize" the differences in grey matter based on a slide-by-slide generation of differences between the aligned baseline and the 12-months follow-up exam and the addition of these differences to the follow-up exam. This method was applied in a CAD system and the results showed that the information extracted by this process contributed to the success of this approach when used in combination with VBM and deep learning techniques.

The first diagnostic goal of this thesis was to differentiate MCI patients who will develop AD among all MCI patients. The study presented a CAD model that achieved an accuracy of 78.7% (80.6% sensitivity and 76.4% specificity) by using the baseline T1-weighted MR images. The other CAD model yielded an accuracy of 97% (sensitivity of 96.8% and specificity of 97.3%), which outperformed the existing studies in the literature. Another diagnostic goal of this thesis was to distinguish age-related cognitive decline from other dementias. The pairwise classification results show that ARCD is completely separable from AD with a success rate of 100% and highly separable from MCI and VD with success rates of 95.4% and 86.30%, respectively.

In summary this work contributes to the literature by proposing a novel pooling method, a longitudinal analysis method, two CAD models for MCI patients based on deep learning techniques and one CAD model for age-related cognitive decline patients using machine learning techniques. This work can be further enlarged by including other type of diseases presenting similar characteristics as AD.

REFERENCES

1.   A. Traboulsee and D. Li. The Role of MRI in the Diagnosis of Multiple Sclerosis. *Advances in Neurology*, 98: 125–146, 2006.

2.   F. Meijer and B. Goraj. Brain MRI in Parkinson's Disease. *Frontiers in Bioscience*, 6: 360–369, 2014.

3.   G. Frisoni, N. Fox, C. Jack, P. Scheltens, and P. Thompson. The Clinical Use of Structural MRI in Alzheimer Disease. *Nature Reviews Neurology*, 2: 67–77, 2010.

4.   N. Mendhiratta, S. Taneja, and A. Rosenkrantz. The Role of MRI in Prostate Cancer Diagnosis and Management. *Future Oncology*, 12: 2431–2443, 2016.

5.   A. Swayampakula, C. Dillis, and J. Abraham. Role of MRI in Screening, Diagnosis and Management of Breast Cancer. *Expert Review of Anticancer Therapy*, 5: 811–817, 2008.

6.   N. Upadhyay and A. Waldman. Conventional MRI Evaluation of Gliomas. *British Journal of Radiology*, 2: 107–111, 2011.

7.   R. Tombropoulos, S. Shiffman, and C. Davidson. A Decision Aid for Diagnosis of Liver Lesions on MRI. *Symposium on Computer Applications in Medical Care (SCAMC), 17th Annual Symposium*, Washington, 439–443, 1993.

8.   W. Schima. MRI of the Pancreas: Tumours and Tumour-Simulating Processes. *Cancer Imaging*, 1: 199–203, 2006.

9.   A. Paciorek, A. Urbanik, A. Brzozowska-Czarnek, J. Paciorek, and I. Herman-Sucharska. The Role of MRI in the Assessment of the Cardiac Anatomy and Function. *Przegląd Lekarski*, 4: 306–313, 2010.

10.   W. Kalender. X-ray Computed Tomography. *Physics in Medicine and Biology*, 13: 29–43, 2006.

11.   W. R. Hendee and E. R. Ritenour. *Medical Imaging Physics*, John Wiley and Sons, Inc., Canada, 2003.

12.   N. Petrick, B. Sahiner, S. G. Armato, A. Bert, L. Correale, S. Delsanto, M. T. Freedman, D. Fryd, D. Gur, L. Hadjiiski, Z. Huo, Y. Jiang, L. Morra, S. Paquerault, V. Raykar, F. Samuelson, R. M. Summers, G. Tourassi, H. Yoshida, B. Zheng, C. Zhou, and H. P. Chan. Evaluation of Computer-aided Detection and Diagnosis Systems. *Medical Physics*, 8: 87–101, 2013.

13.   M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri. The Global Prevalence of Dementia: A Systematic Review and Metaanalysis. *Alzheimer's and Dementia*, 9: 63–75, 2013.

14.   V. T. Cotter. The Burden of Dementia. *American Journal of Managed Care*, 8: 12–19, 2007.

15.   Alzheimer's Disease International. World Alzheimer Report 2015, The Global Economic Impact of Dementia. *Alzheimer's Disease International*, 1: 1–52, 2015.

16.   K. Kuruppuarachchi and T. Lawrence. Carer Burden in Dementia. *The Ceylon Medical Journal*, 1: 44–52, 2006.

17.   C. Qiu, M. Kivipelto, and E. Von Strauss. Epidemiology of Alzheimer's Disease: Occurrence, Determinants, and Strategies Toward Intervention. *Dialogues in Clinical Neuroscience*, 2: 111–128, 2009.

18.   G. Roman, T. Tatemichi, T. Erkinjuntti, J. Cummings, J. Masdeu, J. Garcia, L. Amaducci, J. Orgogozo, A. Brun, and A. Hofman. Vascular Dementia: Diagnostic Criteria for Research Studies: Report of the NINDS-AIREN International Workshop. *Neurology*, 24: 250–260, 1993.

19.   J. Noble, P. Canoll, and L. Honig. Brain Tumor-associated Dementia. *Science of Aging Knowledge Environment*, 34: 2–10, 2005.

20.   A. Futamura, Y. Mori, and M. Kawamura. Diabetes and Dementia. *Brain and Nerve*, 6: 725–732, 2015.

21.   E. Hobik, R. Ihl, and C. Kretschmar. Dementia and Disorders of the Thyroid Gland. *Fortschritte der Neurologie-Psychiatrie*, 9: 330–336, 1994.

22.   A. Korczyn and I. Halperin. Depression and Dementia. *Journal of the Neurological Sciences*, 1: 139–142, 2009.

23.   Y. Moriyama, M. Mimura, M. Kato, and H. Kashima. Primary Alcoholic Dementia and Alcohol-related Dementia. *Psychogeriatrics*, 3: 114–118, 2006.

24.   L. Crews and E. Masliah. Molecular Mechanisms of Neurodegeneration in Alzheimer's Disease. *Human Molecular Genetics*, 19: 12–20, 2010.

25.   Y. Bengio, A. Courville, and P. Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8: 1798–1828, 2013.

26.   C. Reid-Turner, A. Fuggetta, L. Lavazza, and A. Wolf. A Conceptual Basis for Feature Engineering. *Journal of Systems and Software*, 1: 3–15, 1999.

27.   Y. Shim, D. Yang, C. Roe, M. Coats, T. Benzinger, C. Xiong, J. Galvin, N. Cairns, and J. Morris. Pathological Correlates of White Matter Hyperintensities on Magnetic Resonance Imaging. *Dementia and Geriatric Cognitive Disorders*, 1: 92–104, 2015.

28.   W. Klunk, H. Engler, A. Nordberg, Y. Wang, G. Blomqvist, D. P. Holt, M. Bergström, I. Savitcheva, G. F. Huang, S. Estrada, B. Ausén, M. Debnath, J. Barletta, J. C. Price, J. Sandell, B. J. Lopresti, A. Wall, P. Koivisto, G. Antoni, C. Mathis, and B. Långström. Imaging Brain Amyloid in Alzheimer's Disease with Pittsburgh Compound-B. *Annals of Neurology*, 3: 306–319, 2004.

29.   W. E. Klunk. Amyloid Imaging as a Biomarker for Cerebral β-amyloidosis and Risk Prediction for Alzheimer Dementia. *Neurobiology of Aging*, 1: 20–36, 2011.

30. L. Griffanti, M. Jenkinson, S. Suri, E. Zsoldos, A. Mahmood, N. Filippini, C. E. Sexton, A. Topiwala, C. Allan, M. Kivimäki, A. Singh-Manoux, K. Ebmeier, C. Mackay, and G. Zamboni. Classification and Characterization of Periventricular and Deep White Matter Hyperintensities on MRI: A Study in Older Adults. *NeuroImage*, 17: 56–68, 2017.

31. H. G. Schnack, H. E. Hulshoff Pol, W. Baare, M. Viergever, and R. Kahn. Automatic Segmentation of the Ventricular System from MR Images of the Human Brain. *NeuroImage*, 14: 95–104, 2001.

32. B. Magnin, L. Mesrob, S. Kinkingnehun, M. Pelegrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehericy, and H. Benali. Support Vector Machine-based Classification of Alzheimer's Disease from Whole-brain Anatomical MRI. *Neuroradiology*, 51: 73–83, 2009.

33. S. Klöppel, C. Stonnington, C. Chu, B. Draganski, R. Scahill, J. Rohrer, N. Fox, C. Jack, J. Ashburner, and R. Frackowiak. Automatic Classification of MR Scans in Alzheimer's Disease. *Brain*, 131: 681–689, 2008.

34. P. Vemuri, J. Gunter, M. L. Senjem, J. Whitwell, K. Kantarci, D. Knopman, B. F. Boeve, R. Petersen, and C. R. Jack. Alzheimer's Disease Diagnosis in Individual Subjects using Structural MR images: Validation Studies. *NeuroImage*, 39: 1186–1197, 2008.

35. Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos. COMPARE: Classification of Morphological Patterns Using Adaptive Regional Elements. *IEEE Transactions on Medical Imaging*, 26: 93–105, 2007.

36. C. Davatzikos, X. Da, J. B. Toledo, J. Zee, D. A. Wolk, S. X. Xie, Y. Ou, A. Shacklett, P. Parmpi, L. Shaw, and J. Q. Trojanowski. Integration and Relative Value of Biomarkers for Prediction of MCI to AD Progression: Spatial Patterns of Brain Atrophy, Cognitive Scores, APOE Genotype and CSF Biomarkers. *NeuroImage*, 4: 164–173, 2014.

37. C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick. Detection of Prodromal Alzheimer's Disease via Pattern Classification of Magnetic Resonance Imaging. *Neurobiology of Aging*, 29: 514–523, 2008.

38.   Z. Lao, D. Shen, Z. Xue, B. Karacali, S. Resnick, and C. Davatzikos. Morphological Classification of Brains via High-dimensional Shape Transformations and Machine Learning Methods. *NeuroImage*, 1: 46–57, 2004.

39.   Y. Fan, R. E. Gur, R. C. Gur, X. Wu, D. Shen, M. E. Calkins, and C. Davatzikos. Unaffected Family Members and Schizophrenia Patients Share Brain Structure Patterns: A High-Dimensional Pattern Classification Study. *Biological Psychiatry*, 63: 118–124, 2008.

40.   D. Schmitter, A. Roche, B. Maréchal, D. Ribes, A. Abdulkadir, M. Bach-Cuadra, A. Daducci, C. Granziera, S. Klöppel, P. Maeder, R. Meuli, G. Krueger, and ADNI. An Evaluation of Volume-based Morphometry for Prediction of Mild Cognitive Impairment and Alzheimer's Disease. *NeuroImage: Clinical*, 1: 7–17, 2014.

41.   O. Colliot, G. Chetelat, M. Chupin, B. Desgranges, B. Magnin, H. Benali, B. Dubois, L. Garnero, F. Eustache, and S. Lehericy. Discrimination Between Alzheimer Disease, Mild Cognitive Impairment, and Normal Aging by Using Automated Segmentation of the Hippocampus. *Radiology*, 248: 194–201, 2008.

42.   M. Chupin, A. Hammers, R. Liu, O. Colliot, J. Burdett, E. Bardinet, J. S. Duncan, L. Garnero, and L. Lemieux. Automatic Segmentation of the Hippocampus and the Amygdala Driven by Hybrid Constraints: Method and Validation. *NeuroImage*, 46: 749–761, 2009.

43.   M. Chupin, E. Gerardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehericy, H. Benali, L. Garnero, and O. Colliot. Fully Automatic Hippocampus Segmentation and Classification in Alzheimer's Disease and Mild Cognitive Impairment Applied on Data from ADNI. *Hippocampus*, 19: 579–587, 2009.

44.   O. Querbes, F. Aubry, J. Pariente, J. Lotterie, J. Demonet, V. Duret, M. Puel, I. Berry, J. C. Fort, and P. Celsis. Early Diagnosis of Alzheimer's Disease Using Cortical Thickness: Impact of Cognitive Reserve. *Journal of Neurology*, 132: 2036–2047, 2009.

45.   L. Yuan, Y. Wang, P. Thompson, V. Narayan, and J. Ye. Multi-source Feature Learning for Joint Analysis of Incomplete Multiple Heterogeneous Neuroimaging Data. *NeuroImage*, 61: 622–632, 2012.

46. E. Gerardin, G. Chetelat, M. Chupin, R. Cuingnet, B. Desgranges, H. Kim, M. Niethammer, B. Dubois, S. Lehericy, L. Garnero, F. Eustache, and O. Colliot. Multidimensional Classification of Hippocampal Shape Features Discriminates Alzheimer's Disease and Mild Cognitive Impairment from Normal Aging. *NeuroImage*, 47: 1476–1486, 2009.

47. J. McCarthy and E. A. Feigenbaum. In Memoriam: Arthur Samuel: Pioneer in Machine Learning. *AI Magazine*, 3: 10–11, 1990.

48. W. S. McCulloch and W. Pitts. A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, 2: 73–97, 1990.

49. D. Hebb, G. Allport, and E. Kandel. Hebbian Theory. *Current Opinion in Neurobiology*, 5: 1–4, 1949.

50. F. Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 6: 386–408, 1958.

51. C. Cortes and V. Vapnik. Support Vector Machine. *Machine Learning*, 8: 1303–1308, 1995.

52. S. Vieira, W. H. L. Pinaya, and A. Mechelli. Using Deep Learning to Investigate the Neuroimaging Correlates of Psychiatric and Neurological Disorders: Methods and Applications. *Neuroscience and Biobehavioral Reviews*, 74: 58–75, 2017.

53. J. Schmidhuber. Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61: 85–117, 2014.

54. K. Fukushima. Neocognitron: A Hierarchical Neural Network Capable of Visual Pattern Recognition. *Neural Networks*, 2: 119–130, 1988.

55. S. Hochreiter and J. J. Schmidhuber. Long Short-term Memory. *Neural Computation*, 8: 1–32, 1997.

56.   G. Hinton. Deep Belief Networks. *Scholarpedia*, 4: 97–108, 2009.

57.   R. Salakhutdinov and G. Hinton. Deep Boltzmann Machines. *International Conference on Artificial Intelligence and Statistics (AISTATS), 12th International Conference*, Florida, 448–455, 2009.

58.   A. Coates, A. Arbor, and A. Ng. An Analysis of Single-Layer Networks in Unsupervised Feature Learning. *International Conference on Artificial Intelligence and Statistics (AISTATS), 14th International Conference*, Fort Lauderdale, 215–223, 2011.

59.   G. Hinton. A Practical Guide to Training Restricted Boltzmann Machines. *Momentum*, 1: 926–927, 2010.

60.   M. Carreira-Perpinan and G. Hinton. On Contrastive Divergence Learning. *International Conference on Artificial Intelligence and Statistics (AISTATS), 11th International Conference*, Barbados, 17–21, 2005.

61.   A. Fischer and C. Igel. Training Restricted Boltzmann Machines: An Introduction. *Pattern Recognition*, 1: 25–39, 2014.

62.   Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. Greedy Layer-Wise Training of Deep Networks. *Advances in Neural Information Processing Systems*, 1: 153, 2007.

63.   J. Ashburner and K. J. Friston. Voxel-based Morphometry The Methods. *NeuroImage*, 11: 805–821, 2000.

64.   J. L. Whitwell. Voxel-based Morphometry: An Automated Technique for Assessing Structural Changes in the Brain. *The Journal of Neuroscience*, 31: 9661–9664, 2009.

65.   J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster. A Probabilistic Atlas of the Human Brain: Theory and Rationale for its Development. The International Consortium for Brain Mapping (ICBM). *NeuroImage*, 2: 89–101, 1995.

66.   J. Ashburner and K. J. Friston. Unified Segmentation. *NeuroImage*, 3: 839–851,

2005.

67.   C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. G. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI Methods. *Journal of Magnetic Resonance Imaging*, 27: 685–91, 2008.

68.   L. Khedher, J. Ramirez, J. M. Gorriz, and A. Brahim. Automatic Classification of Segmented MRI Data Combining Independent Component Analysis and Support Vector Machines. *Studies in Health Technology and Informatics*, 207: 271–279, 2014.

69.   R. Cuingnet, E. Gérardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, O. Colliot, and E. Gerardin. Automatic Classification of Patients with Alzheimer's Disease from Structural MRI: A Comparison of Ten Methods Using the ADNI Database. *NeuroImage*, 56: 766–81, 2010.

70.   H. Il Suk, S. W. Lee, and D. Shen. Hierarchical Feature Representation and Multimodal Fusion with Deep Learning for AD/MCI Diagnosis. *NeuroImage*, 101: 569–582, 2014.

71.   H. Il Suk and D. Shen. Deep Learning Based Feature Representation for AD/MCI Classification. *Medical Image Computing and Computer-Assisted Intervention (MICCAI), 16th International Conference*, Nagoya, 583–590, 2013.

72.   D. Zhang and D. Shen. Multi-modal Multi-task Learning for Joint Prediction of Multiple Regression and Classification Variables in Alzheimer's Disease. *NeuroImage*, 59: 895–907, 2012.

73.   X. Zhu, H. Il Suk, and D. Shen. A Novel Matrix Similarity Based Loss Function for Joint Regression and Classification in AD Diagnosis. *NeuroImage*, 100: 91–105, 2014.

74.   E. Veronese, U. Castellani, D. Peruzzo, M. Bellani, and P. Brambilla. Machine Learning Approaches: From Theory to Application in Schizophrenia. *Computational and Mathematical Methods in Medicine*, 2: 1–12, 2013.

75.   H. Mishan-Shamay, G. Doniger, E. Chalom, E. Simon, and R. Unger. A Machine-learning Approach for Integration of Computerized Cognitive Data in the Neuropsychological Assessment of Older Adults. *Alzheimer's and Dementia: The Journal of the Alzheimer's Association*, 4: 635–636, 2015.

76.   A. Järvelin and M. Juhola. Comparison of Machine Learning Methods for Classifying Aphasic and Non-aphasic Speakers. *Computer Methods and Programs in Biomedicine*, 104: 349–357, 2011.

77.   A. Jarvelin, M. Juhola, and M. Laine. A Neural Network Model for the Simulation of Word Production Errors of Finnish Nouns. *International Journal of Neural Systems*, 16: 241–254, 2006.

78.   M. J. Wu, I. C. Passos, I. E. Bauer, L. Lavagnino, B. Cao, G. B. Zunta-Soares, F. Kapczinski, B. Mwangi, and J. C. Soares. Individualized Identification of Euthymic Bipolar Disorder using the Cambridge Neuropsychological Test Automated Battery (CANTAB) and Machine Learning. *Journal of Affective Disorders*, 192: 219–225, 2016.

79.   S. H. Hojjati, A. Ebrahimzadeh, A. Khazaee, and A. Babajani-Feremi. Predicting Conversion from MCI to AD Using Resting-state fMRI, Graph Theoretical Approach and SVM. *Journal of Neuroscience Methods*, 282: 69–80, 2017.

80.   J. Ye, M. Farnum, E. Yang, R. Verbeeck, V. Lobanov, N. Raghavan, G. Novak, A. DiBernardo, and V. A. Narayan. Sparse Learning and Stability Selection for Predicting MCI to AD Conversion Using Baseline ADNI Data. *BMC Neurology*, 1: 46, 2012.

81.   A. Gupta, M. S. Ayhan, and A. S. Maida. Evaluation of Autoencoders for Bases to Represent Neuroimaging Data. *Workshop on Machine Learning and Interpretation in NeuroImaging (MLINI), 3rd NIPS 2013 Workshop*, Nevada, 7–12, 2013.

82.   A. Gupta, M. S. Ayhan, and A. S. Maida. Natural Image Bases to Represent Neuroimaging Data. *International Conference on Machine Learning (ICML), 30th International Conference*, Georgia, 28–29, 2013.

83.   Minhas, Khanum, Riaz, Alvi, and Khan. A Non Parametric Approach for Mild Cognitive Impairment to AD Conversion Prediction: Results on Longitudinal Data. *IEEE Journal of Biomedical and Health Informatics*, 99: 1–2, 2016.

84.   D. P. Devanand, X. Liu, M. H. Tabert, G. Pradhaban, K. Cuasay, K. Bell, M. J. de Leon, R. L. Doty, Y. Stern, and G. H. Pelton. Combining Early Markers Strongly Predicts Conversion from Mild Cognitive Impairment to Alzheimer's Disease. *Biological Psychiatry*, 10: 871–879, 2008.

85.   A. A. Willette, V. D. Calhoun, J. M. Egan, and D. Kapogiannis. Prognostic Classification of Mild Cognitive Impairment and Alzheimer's Disease: MRI Independent Component Analysis. *Psychiatry Research - Neuroimaging*, 2: 81–88, 2014.

86.   B. A. Ardekani, E. Bermudez, A. M. Mubeen, A. H. Bachman, and Alzheimer's Disease Neuroimaging Initiative. Prediction of Incipient Alzheimer's Disease Dementia in Patients with Mild Cognitive Impairment. *Journal of Alzheimer's Disease*, 1: 269–281, 2016.

87.   F. Peters, S. Villeneuve, and S. Belleville. Predicting Progression to Dementia in Elderly Subjects with Mild Cognitive Impairment using Both Cognitive and Neuroimaging Predictors. *Journal of Alzheimer's Disease*, 2: 307–318, 2014.

88.   F. Liu, H.-I. Suk, C.-Y. Wee, H. Chen, and D. Shen. High-Order Graph Matching Based Feature Selection for Alzheimer's Disease Identification. *Medical Image Computing and Computer-Assisted Intervention (MICCAI), 16th International Conference*, Nagoya, 311–318, 2013.

89.   X. Zhu, H.-I. Suk, and D. Shen. Multi-modality Canonical Feature Selection for Alzheimer's Disease Diagnosis. *Medical Image Computing and Computer-Assisted Intervention (MICCAI), 17th International Conference*, Boston, 162–169, 2014.

90.   M. Liu, D. Zhang, and D. Shen. Identifying Informative Imaging Biomarkers via Tree Structured Sparse Learning for AD Diagnosis. *Neuroinformatics*, 12: 381–394, 2014.

91.   A. Ng. Unsupervised Feature Learning and Deep Learning, UFLDL's Homepage, http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial, [retrieved 01 April 2017].

92.   A. Payan and G. Montana. Predicting Alzheimer's Disease: A Neuroimaging Study with 3D Convolutional Neural Networks. 5: 355–362, 2015.

93.   Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 4: 541–551, 1989.

94.   Z. Shi, Y. Ye, and Y. Wu. Rank-based Pooling for Deep Convolutional Neural Networks. *Neural Networks*, 83: 21–31, 2016.

95.   Y. Liu, Y. M. Zhang, X. Y. Zhang, and C. L. Liu. Adaptive Spatial Pooling for Image Classification. *Pattern Recognition*, 55: 58–67, 2016.

96.   M. J. Er, Y. Zhang, N. Wang, and M. Pratama. Attention Pooling-based Convolutional Neural Network for Sentence Modelling. *Information Sciences*, 373: 1339–1351, 2016.

97.   J. Feng, B. Ni, Q. Tian, and S. Yan. Geometric LP-norm Feature Pooling for Image Classification. *IEEE Computer Vision and Pattern Recognition (CVPR), 24th International Conference*, Colorado, 2697–2704, 2011.

98.   B. Graham. Fractional Max-Pooling. *International Conference on Learning Representations (ICLR), 2nd International Conference*, Banff, 1–10, 2014.

99.   M. D. Zeiler and R. Fergus. Stochastic Pooling for Regularization of Deep Convolutional Neural Networks. *International Conference on Learning Representations (ICLR), 1st International Conference*, Scottsdale, 1–9, 2013.

100.   W. Xiong, L. Zhang, B. Du, and D. Tao. Combining Local and Global: Rich and Robust Feature Pooling for Visual Recognition. *Pattern Recognition*, 62: 225–235, 2017.

101.   Y. Huang, Z. Wu, L. Wang, and C. Song. Multiple Spatial Pooling for Visual Object Recognition. *Neurocomputing*, 129: 225–231, 2014.

102.   A. J. Mitchell and M. Shiri-Feshki. Rate of Progression of Mild Cognitive Impairment to Dementia - Meta-analysis of 41 Robust Inception Cohort Studies. *Acta Psychiatrica Scandinavica*, 4: 252–265, 2009.

103.   R. Tarawneh and D. M. Holtzman. The Clinical Problem of Symptomatic Alzheimer Disease and Mild Cognitive Impairment. *Cold Spring Harbor Perspectives in Medicine*, 5: 45–58, 2012.

104.   S. Risacher and A. Saykin. Neuroimaging Biomarkers of Neurodegenerative Diseases and Dementia. *Seminars in Neurology*, 4: 386–416, 2013.

105.   K. Blennow and H. Zetterberg. Cerebrospinal Fluid Biomarkers for Alzheimer's Disease. *Journal of Alzheimer's Disease*, 2: 413–417, 2009.

106.   S. Kiddle, C. Steves, M. Mehta, A. Simmons, X. Xu, S. Newhouse, M. Sattlecker, N. Ashton, C. Bazenet, R. Killick, J. Adnan, E. Westman, S. Nelson, H. Soininen, I. Kloszewska, P. Mecocci, M. Tsolaki, B. Vellas, C. Curtis, G. Breen, S. C. R. Williams, S. Lovestone, T. D. Spector, and R. J. B. Dobson. Plasma Protein Biomarkers of Alzheimer's Disease Endophenotypes in Asymptomatic Older Twins: Early Cognitive Decline and Regional Brain Volumes. *Translational Psychiatry*, 5: 584–592, 2015.

107.   B. Mahendra. Some Ethical Issues in Dementia Research. *Journal of Medical Ethics*, 1: 29–31, 1984.

108.   E. Howe. Ethical Issues in Diagnosing and Treating Alzheimer Disease. *Psychiatry*, 5: 43–52, 2006.

109.   A. Baird, S. Westwood, and S. Lovestone. Blood-based Proteomic Biomarkers of

Alzheimer's Disease Pathology. *Frontiers in Neurology*, 6: 236–242, 2015.

110.   A. Del Sole, S. Malaspina, and A. Magenta-Biasina. Magnetic Resonance Imaging and Positron Emission Tomography in the Diagnosis of Neurodegenerative Dementias. *Functional Neurology*, 4: 205–215, 2016.

111.   K. C. Fraser, J. a. Meltzer, N. L. Graham, C. Leonard, G. Hirst, S. E. Black, and E. Rochon. Automated Classification of Primary Progressive Aphasia Subtypes from Narrative Speech Transcripts. *Cortex*, 55: 43–60, 2014.

112.   P. Garrard, V. Rentoumi, B. Gesierich, B. Miller, and M. L. Gorno-Tempini. Machine Learning Approaches to Diagnosis and Laterality Effects in Semantic Dementia Discourse. *Cortex*, 55: 122–129, 2014.

113.   D. Clark, P. Kapur, D. Geldmacher, J. Brockington, L. Harrell, T. DeRamus, P. Blanton, K. Lokken, A. Nicholas, and D. C. Marson. Latent Information in Fluency Lists Predicts Functional Decline in Persons at Risk for Alzheimer Disease. *Cortex*, 55: 202–218, 2014.

114.   D. R. Roalf, P. J. Moberg, S. X. Xie, D. A. Wolk, S. T. Moelter, and S. E. Arnold. Comparative Accuracies of Two Common Screening Instruments for Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Healthy Aging. *Alzheimer's and Dementia*, 5: 529–537, 2013.

115.   J. Maroco, D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, and A. De Mendonça. Data Mining Methods in the Prediction of Dementia: A Real-data Comparison of the Accuracy, Sensitivity and Specificity of Linear Discriminant Analysis, Logistic Regression, Neural Networks, Support Vector Machines, Classification Trees and Random Forests. *BMC Research Notes*, 299: 12–19, 2011.

116.   D. Silva, M. Guerreiro, J. Maroco, I. Santana, A. Rodrigues, J. Bravo Marques, and A. De Mendonça. Comparison of Four Verbal Memory Tests for the Diagnosis and Predictive Value of Mild Cognitive Impairment. *Dementia and Geriatric Cognitive Disorders Extra*, 1: 120–131, 2012.

117.   D. L. Green, L. Payne, R. Polikar, P. J. Moberg, D. A. Wolk, and J. Kounios. A Candidate ERP Biomarker of Prodromal Alzheimer's Disease. *Brain Research*, 1624: 390–397, 2015.

118.   A. Nazeri, H. Ganjgahi, T. Roostaei, T. Nichols, and M. Zarei. Imaging Proteomics for Diagnosis, Monitoring and Prediction of Alzheimer's Disease. *NeuroImage*, 2: 657–665, 2014.

119.   G. Douaud, R. a L. Menke, A. Gass, A. U. Monsch, A. Rao, B. Whitcher, G. Zamboni, P. M. Matthews, M. Sollberger, and S. Smith. Brain Microstructure Reveals Early Abnormalities More than Two Years Prior to Clinical Progression from Mild Cognitive Impairment to Alzheimer's Disease. *The Journal of Neuroscience*, 5: 2147–2155, 2013.

120.   J. Dukart, F. Sambataro, and A. Bertolino. Accurate Prediction of Conversion to Alzheimer's Disease using Imaging, Genetic, and Neuropsychological Biomarkers. *Journal of Alzheimer's Disease*, 4: 1143–1159, 2015.

121.   A. Ortiz, J. Munilla, J. M. Górriz, and J. Ramírez. Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer's Disease. *International Journal of Neural Systems*, 7: 165–185, 2016.

122.   I. O. Korolev, L. L. Symonds, A. C. Bozoki, and A. D. N. Initiative. Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification. *Plos One*, 2: 1–10, 2016.

123.   H.-I. Suk, S.-W. Lee, and D. Shen. Latent Feature Representation with Stacked Auto-encoder for AD/MCI Diagnosis. *Brain Structure and Function*, 220: 841–859, 2015.

124.   R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, and S. Ji. Deep Learning based Imaging Data Completion for Improved Brain Disease Diagnosis. *Medical Image Computing and Computer-Assisted Intervention (MICCAI), 17th International Conference*, Boston, 305–312, 2014.

125. S. M. Plis, D. R. Hjelm, R. Slakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, H. Johnson, J. Paulsen, J. Turner, and V. D. Calhoun. Deep Learning for Neuroimaging: A Validation Study. *Frontiers in Neuroscience*, 8: 1–11, 2014.

126. J. C. Morris, M. Storandt, J. P. Miller, D. W. McKeel, J. L. Price, E. H. Rubin, and L. Berg. Mild Cognitive Impairment Represents Early-stage Alzheimer Disease. *Archives of Neurology*, 3: 397–405, 2001.

127. D. J. Gracey and R. G. Morris. Neuropsychological Assessment in Dementia. *Psychiatry*, 12: 498–502, 2007.

128. G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub, and C. H. Phelps. The Diagnosis of Dementia Due to Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease. *Alzheimer's and Dementia*, 3: 263–269, 2011.

129. M. S. Albert, S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen, P. J. Snyder, M. C. Carrillo, B. Thies, and C. H. Phelps. The Diagnosis of Mild Cognitive Impairment due to Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease. *Alzheimer's and Dementia*, 3: 270–279, 2011.

130. A. L. Benton, A. B. Sivan, K. Hamsher, N. R. Varney, and O. Spreen. *Contribution to Neuropsychological Assessment*, Oxford University Press, New York, 1983.

131. K. I. Shulman. Clock-drawing: Is it the Ideal Cognitive Screening Test? *International Journal of Geriatric Psychiatry*, 6: 548–561, 2000.

132. J. R. Stroop. Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology*, 6: 643–662, 1935.

133. E. F. Kaplan, H. Goodglass, and S. Weintraub. *The Boston Naming Test*, Lea and Febiger, Inc., Philadelphia, 1983.

134. Ö. Öktem. A Verbal Test of Memory Processes: A Preliminary Study. *Nöropsikiyatri Arşivi*, 4: 196–206, 1992.

135. W. H. Kruskal and W. A. Wallis. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 260: 583–621, 1952.

136. R. Bender, A. Ziegler, and S. Lange. Analysis of Variance. *Deutsche Medizinische Wochenschrift*, 132: 57–60, 2007.

137. H. B. Mann and D. R. Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 1: 50–60, 1947.

138. H. Mann and D. Whitney. On a Test of Whether One of Two Random Variables is Stochastically Larger Than the Other. *Annals of Mathematical Statistics*, 1: 50–60, 1947.

139. W. J. Conover. Practical Nonparametric Statistics. *The Statistician*, 22: 309–314, 1971.

140. Q. McNemar. Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages. *Psychometrika*, 2: 153–157, 1947.

141. S. Labovitz. Criteria for Selecting a Significance Level: A Note on the Sacredness of .05. *The American Sociologist*, 3: 220–222, 1968.

142. F. Er, D. Goularas, and B. Ormeci. A Novel Convolutional Neural Network Model Based on Voxel-based Morphometry of Imaging Data in Predicting the Prognosis of Patients with Mild Cognitive Impairment. *Journal of Neurological Sciences (Turkish)*, 1: 52–69, 2017.

143. F. Er, P. Iscen, S. Sahin, N. Cinar, S. Karsidag, and D. Goularas. Distinguishing Age-

related Cognitive Decline from Dementias: A Study Based on Machine Learning Algorithms. *Journal of Clinical Neuroscience*, 4: 52–64, 2017.

144.   I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene Selection for Cancer Classification Using Support Vector Machines. *Machine Learning*, 3: 389–422, 2002.

145.   N. Peddakam, S. Susaria, and A. S. Reddy. Multilayer Perceptron Algorithm (XOR) using Backpropagation. *International Journal of Advanced Research in Computer Science and Software Engineering*, 15: 319–322, 2015.

146.   R. Kohavi. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence (IJCAI), 14th International Conference*, Quebec, 1137–1143, 1995.

147.   M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA Data Mining Software: an Update. *ACM SIGKDD Explorations Newsletter*, 11: 10–18, 2009.

148.   C. Bortolon, A. Louche, M. C. Gely-Nargeot, and S. Raffard. Do Patients Suffering from Alzheimer's Disease Present an Own-age Bias in Face Recognition? *Experimental Gerontology*, 70: 46–53, 2015.

149.   D. J. Gracey and R. G. Morris. Neuropsychological Assessment in Dementia. *Psychiatry*, 6: 498–502, 2007.

150.   Y. Saeys, I. Inza, and P. Larrañaga. A Review of Feature Selection Techniques in Bioinformatics. *Bioinformatics*, 23: 2507–2517, 2007.

151.   F. Barwick, P. Arnett, and S. Slobounov. EEG Correlates of Fatigue During Administration of a Neuropsychological Test Battery. *Clinical Neurophysiology*, 123: 278–284, 2012.

152.   Z. Xu, X. Shen, and W. Pan. Longitudinal Analysis is More Powerful than Cross-sectional Analysis in Detecting Genetic Association with Neuroimaging Phenotypes. *Plos*

One, 8: 1–10, 2014.

153.   C. Morgan, J. Pearson, and G. Fuller. Lumbar Punctures and Cerebrospinal Fluid Analysis. *Medicine*, 8: 504–507, 2016.

154.   R. W. Evans. Complications of Lumbar Puncture. *Neurologic Clinics*, 1: 83–105, 1998.

155.   N. Sharma and A. N. Singh. Exploring Biomarkers for Alzheimer's Disease. *Journal of Clinical and Diagnostic Research*, 7: 1–6, 2016.

156.   G. Chetelat and J. C. Baron. Early Diagnosis of Alzheimer's Disease: Contribution of Structural Neuroimaging. *NeuroImage*, 2: 525–541, 2003.

157.   C. H. Samer, K. Rishi, and C. Rowen. Image Recognition Using Convolutional Neural Networks. *Cadence Whitepaper*, 7: 1–12, 2015.