

11-13-2020

A Multimodal Neuroimaging Approach for Classification and Prediction of Alzheimer's Disease Using Machine Learning

Parisa Forouzannezhad
pforo003@fiu.edu

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>



Part of the [Other Electrical and Computer Engineering Commons](#)

Recommended Citation

Forouzannezhad, Parisa, "A Multimodal Neuroimaging Approach for Classification and Prediction of Alzheimer's Disease Using Machine Learning" (2020). *FIU Electronic Theses and Dissertations*. 4584.
<https://digitalcommons.fiu.edu/etd/4584>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

A MULTIMODAL NEUROIMAGING APPROACH FOR CLASSIFICATION AND
PREDICTION OF ALZHEIMER'S DISEASE USING MACHINE LEARNING

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL AND COMPUTER ENGINEERING

by

Parisa Forouzannezhad

2020

To: Dean John L. Volakis
College of Engineering and Computing

This dissertation, written by Parisa Forouzannezhad, and entitled A Multimodal Neuroimaging Approach for Classification and Prediction of Alzheimer’s Disease using Machine Learning, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Jean Andrian

Armando Barreto

Mercedes Cabrerizo

Naphtali Rishe

Malek Adjouadi, Major Professor

Date of Defense: November 13, 2020

The dissertation of Parisa Forouzannezhad is approved.

Dean John L. Volakis
College of Engineering and Computing

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2020

© Copyright 2020 by Parisa Forouzannezhad
All rights reserved.

DEDICATION

I dedicate this dissertation to my loving parents. Without their love, support, and patience, it would have been impossible to complete this research.

ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to my major advisor, Dr. Malek Adjouadi for his encouragement, patience, and guidance in my research and for his kindness and consideration throughout the past four years that I have worked with him at the Center for Advanced Technology and Education-CATE. With his continuous support and invaluable academic advice, I was able to get over all the difficulties, finish my Ph.D., and realize my goal.

Special thanks to all the kind persons that helped me through the past four years. I am grateful to all my committee members Dr. Jean Andrian, Dr. Armando Barreto, Dr. Naphtali Rische, and Dr. Mercedes Cabrerizo for their professional support and assistance. Many thanks go to Dr. Ranjan Duara from Mount Sinai Medical Center and Dr. David Loewenstein from the University of Miami for their valuable suggestions in all clinical and neuropsychological aspects of the Alzheimer's disease.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceu-

tical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This research is supported by the National Science Foundation under grants: CNS-1920182, CNS-1532061, CNS-1338922, CNS-2018611, and CNS-1551221, and by the National Institutes of Health through NIA/NIH grants 1R01AG055638-01A1, 5R01AG061106-02, 5R01AG047649-05, and the 1P30AG066506-01 with the 1Florida Alzheimer's Disease Research Center (ADRC). Our thanks also go to the Ware Foundation.

ABSTRACT OF THE DISSERTATION
A MULTIMODAL NEUROIMAGING APPROACH FOR CLASSIFICATION AND
PREDICTION OF ALZHEIMER'S DISEASE USING MACHINE LEARNING

by

Parisa Forouzannezhad

Florida International University, 2020

Miami, Florida

Professor Malek Adjouadi, Major Professor

Alzheimer's disease (AD) is one of the most common neurodegenerative disorders among the elderly population. It is progressive, irreversible in nature and is considered the main cause of dementia. AD has become a world health problem affecting developed and developing nations alike, with the number of diagnosed AD patients increasing rather dramatically as both the life span of humans and the earth's population continue to increase. Therefore, AD diagnosis in its earliest manifestations, preferably at the presymptomatic stage is critical for the timely planning of treatment and therapeutic interventions.

We introduce new machine learning algorithms to detect and predict the Alzheimer's disease in the early phase to include the presymptomatic stage where no manifestation of cognitive decline is yet apparent. An investigation is carried out in search of optimal feature selection methods on different machine learning platforms with the intent to address the challenging classification and regression analysis. This research endeavor introduces three machine learning platforms that are based on (1) deep neural network, (2) support vector machine (SVM), and (3) Gaussian-based model classifiers all optimized in order to delineate the different stages of the disease as well as a regression framework to predict future cognitive scores as means to gauge disease progression, which could play an important role in pre- and post-treatment evaluations. The input data to these machine learning architectures included magnetic resonance imaging (MRI), positron emission tomogra-

phy (PET), the metabolic fluorodeoxyglucose (FDG)-PET, cognitive scores, cerebrospinal fluid (CSF), and the apolipoprotein E4 (APOE4) gene. An investigation is carried out on the transition phases of AD through regression analysis by predicting cognitive tests including Alzheimer's disease assessment scale cognitive subscale (ADAS-Cog), Mini-mental state examination (MMSE), and Rey's auditory verbal learning (RAVLT) that have been designed and used as important criteria to evaluate cognitive status of AD patients. We formulated the prediction of disease progression as a multimodal multitask regression problem across six time points in a 4-year longitudinal study.

Major findings of this work reveal that for binary classification, the highest accuracy of 84% for delineating the challenging group of early mild cognitively impaired individuals (EMCI) from the cognitively normal (CN) group is obtained. With multiclass classification using deep neural network methodology, especially when early and late MCI (EMCI and LMCI) groups are included, the accuracy does not exceed 70%, which clearly explains the many nuances in the transition phases of the disease, especially in the early stages. Moreover, the episodic tests like RVALT as used in this study were shown to be effective for selecting the at-risk groups. MRI morphometry was found to be the most sensitive biomarker to predict disease conversion and observed that parietal and prefrontal cortices are also associated with episodic memory in addition to the temporal lobe. Although adding the modalities of FDG-PET, CSF, and APOE allele gene improved the prediction error significantly at 4 time points, multimodal neuroimaging does not statistically enhance the prediction performance at some time points due to the inherent challenge of missing data. It is clear that for longitudinal studies of such duration (4-year), beyond the variability and interrelatedness of features, the missing data challenge remains the most difficult to overcome.

TABLE OF CONTENTS

CHAPTER	PAGE
1. INTRODUCTION	1
2. A DEEP NEURAL NETWORK APPROACH FOR EARLY DETECTION OF MILD COGNITIVE IMPAIRMENT USING MULTIPLE FEATURES	13
2.1 Goal	13
2.2 Materials and Methods	14
2.2.1 Data Acquisition	14
2.2.2 MRI Processing	15
2.2.3 AV-45 PET Processing	16
2.2.4 Statistical Analysis	18
2.2.5 Feature Normalization	19
2.2.6 Deep Neural Network	19
2.3 Results and Discussion	21
3. EARLY DIAGNOSIS OF MILD COGNITIVE IMPAIRMENT USING SUPPORT VECTOR MACHINE	28
3.1 Goal	28
3.2 Materials and Methods	28
3.2.1 Data Acquisition	28
3.2.2 Feature Selection	29
3.2.3 SVM model	32
3.3 Results	33
4. A GAUSSIAN-BASED MODEL FOR EARLY DIAGNOSIS OF MILD COGNITIVE IMPAIRMENT USING MULTIMODAL NEUROIMAGING	37
4.1 Goal	37
4.2 Materials and Methods	37
4.2.1 Data Acquisition	37
4.3 Gaussian Process	38
4.3.1 GP Likelihood	39
4.3.2 GP Prior	40
4.3.3 Marginal likelihood	40
4.3.4 Posterior distribution	41
4.4 Results	42
4.5 Discussion	51
5. PREDICTION OF CONVERSION FROM NORMAL COGNITION AND MILD COGNITIVE IMPAIRMENT TO ALZHEIMER'S DISEASE	54
5.1 Goal	54
5.2 Material and Methods	55

5.2.1	Participants and Data Acquisition	55
5.2.2	Cognitive tests	56
5.2.3	Neuroimaging	58
5.2.4	Fused Gaussian Sparse Group Lasso (FGSGL)	59
5.2.5	Fusion Algorithm	64
5.3	Results	67
5.4	Discussion	77
6.	CONCLUSION	81
	REFERENCES	86
	VITA	104

LIST OF TABLES

TABLE		PAGE
2.1	Demographic information of the participants	15
2.2	Neuropsychological information of the Participants (PF: Percent-Forgetting, Im: Immediate)	15
2.3	Accuracy (Acc), Sensitivity (Sen), and Specificity (Spe) of the DNN classifier for different pairs of binary classification. NTS represents the Neuropsychological Test Scores of RAVLT, ECogT, and MoCA. The values are represented as a percent)	22
2.4	Accuracy (Acc), Sensitivity (Sen), and Specificity (Spe) of the ANN classifier	23
2.5	Comparing the DNN classifier with ANN considering the overall accuracy (Acc), accuracy of CN detection (Ac1), accuracy of EMCI detection (Ac2), accuracy of LMCI detection (Ac3), and accuracy of AD detection (Ac4) for different groups of one vs. all classification (all the values are in percent and rounded to the nearest integer)	25
2.6	Accuracy (Acc) of the DNN classifier comparing to the previous works	26
3.1	Accuracy (Acc), Sensitivity (Sen), and Specificity (Spe) of the SVM classifier for different pairs of binary classification (values are represented as percent)	34
3.2	Accuracy (Acc) of the proposed method compared to the previous studies of EMCI and LMCI	36
4.1	Performance comparison of the proposed method for 6 binary groups	43
4.2	The first eight most important features selected by the feature selection along with the P-value. (lh: left hemisphere, rh: right hemisphere)	46
4.3	Performance Comparison of the Gaussian classifier with SVM using the same kernel and the same feature including the MRI, PET, and DI. (RBF: radial basis function kernel, and L-K: linear kernel)	50
4.4	Computation time in second for SVM and GP using linear and RBF kernels (Variables are represented in seconds)	50
4.5	Accuracy (Acc), sensitivity (Sen), and specificity (Spe) of the Gaussian classifier comparing to the previous works. (dMRI: diffusion magnetic resonance imaging, fMRI: functional magnetic resonance imaging, and DI: demographic information)	53
5.1	Participants Demographic information at baseline. Values are represented as mean (standard deviation)	56
5.2	Participants Demographic information at baseline. Values are represented as mean (standard deviation)	56

5.3	Number of observations for different modalities at different time points of T1-T48	59
5.4	Prediction performance in terms of RMSE for the 4-year time window based on multimodal neuroimaging. Values are represented as mean (standard deviation)	67
5.5	Prediction performance of MMSE and ADAS for the 4-year time window based on multimodal neuroimaging for different groups of subjects separately	70
5.6	Results of our proposed method compared to lasso, ridge, temporal group lasso (TGL), fused group lasso (FGL), and convex fused sparse group lasso (cFSGL) in terms of averaged RMSE (standard deviation) using MRI data to predict MMSE and ADAS. Superscript symbol of * indicates that the marked method significantly outperformed the others at that time point	77

LIST OF FIGURES

FIGURE		PAGE
2.1	Overall diagram of the proposed DNN Classifier.	17
2.2	ROC curve and area under the curve (AUC) of CN vs. EMCI, LMCI, and AD with and without using neuropsychological test scores for the proposed DNN.	23
2.3	ROC curve and AUC values of CN vs. EMCI and CN vs. AD with and without using neuropsychological test scores for both ANN and DNN.	24
2.4	Importance of different neuropsychological test scores for three groups of CN vs. EMCI, CN vs. LMCI, and CN vs. AD classifications.	27
3.1	Diagram of the preprocessing step and the proposed classification.	29
3.2	ROC curve and AUC of three classifications of CN vs. EMCI, CN vs. LMCI, and CN vs. AD using NTS and without using NTS.	35
3.3	Impact of neuropsychological test scores (NTS) used in this study for three classifications of CN vs. EMCI, CN vs. LMCI, and CN vs. AD separately. (PF: Percent-Forgetting, Im: Immediate)	35
4.1	Neuroimaging feature extraction and overall diagram of the proposed method.	38
4.2	The logistic sigmoid function used in the proposed GP learning is plotted based on the latent variable z and class labels of y	39
4.3	Classification results with 95% confidence interval for different modalities for the binary classifications of (a) CN vs. EMCI, (b) CN vs. LMCI, (c) CN vs. AD, (d) EMCI vs. LMCI, (e) EMCI vs. AD, and (f) LMCI vs. AD based on results of Table 4.1.	44
4.4	Receiver Operating Characteristics (ROC) curve and Area Under the Curve (AUC).	45
4.5	Importance of the features listed in Table 4.2 for the most challenging CN vs. EMCI classification.	47
4.6	Boxplot of different features for (a)-(d): CN vs. EMCI, (e)-(h): CN vs. LMCI, and (i)-(l): CN vs. AD.	49
4.7	Classification results with 95% confidence interval for combination of different feature selection using SVM and GP classifiers for the most important classifications of (a) CN vs. EMCI, (b) CN vs. LMCI, and (c) CN vs. AD. (MRMR: minimum redundancy maximum relevance, RF-RFE: random forest recursive feature elimination)	52

5.1	Average changes of cognitive scores for the 5 different groups of CNs, CNc, MCIs, MCIC, and AD across the 6 time points for method for (a) MMSE (b), ADAS (c), RAVLT-Im, and (d) RAVLT-PF.	57
5.2	Overall view of the proposed framework to predict neuropsychological test scores. The portion highlighted in white in the predicted \hat{y}_i^j vectors signify the proportion of missing data.	66
5.3	Scatter plot of predicted MMSE score vs. actual MMSE score along with the correlation coefficient at 6 time points. The orange line is regression line and the dotted gray line is the reference of perfect correlation.	68
5.4	Scatter plot of predicted ADAS score vs. actual ADAS score along with the correlation coefficient at 6 time points. The orange line is regression line and the dotted gray line is the reference of perfect correlation.	69
5.5	Box plot of prediction MMSE score vs. actual MMSE score using the test set based on multimodal neuroimaging for all groups of subjects separately and combined	71
5.6	ROC curves and AUC values of group classification of CNs vs. CNc, MCIs vs. MCIC, CN vs. MCI (includes both groups of stables and converters), and CN vs. AD for cognitive scores of MMSE and ADAS.	73
5.7	ROC curves and AUC values of group classification of CNs vs. CNc, MCIs vs. MCIC, CN vs. MCI (includes both groups of stables and converters), and CN vs. AD for cognitive scores of RAVLT-Im and RAVL-PF.	74
5.8	Bar plot of RMSE for different combination of modalities at different time points in 4-year time window for the cognitive tests of (a) MMSE, (b) ADAS, (c) RAVLT-Im, and (d) RAVLT-PF.	76

CHAPTER 1

INTRODUCTION

Alzheimer's disease (AD) is one of the most common neurodegenerative disorders among the elderly population. It is progressive, irreversible in nature and is considered the main cause of dementia. Therefore, AD diagnosis in its earliest manifestations is critical for the timely planning of treatment of this healthcare challenge. AD has become a world health problem affecting developed and developing nations alike, and the number of diagnosed AD patients is, unfortunately, increasing rather dramatically as both the life span of humans and the earth's population continue to increase. According to the 2018 report by the Alzheimer's Association, 5.7 million Americans are living with Alzheimer's and by 2050 this number may reach 13.8 million [1].

Numerous studies over the recent years have confirmed that AD can be diagnosed by clinical procedures in 90 percent of cases or higher; however, by the time that the AD stage is diagnosed in patients, they might already lost a substantial part of their mental function, and given the irreversible nature of the disease, the chance for early intervention and the potential for slowing its progression become futile. Current studies confirmed that 10-15% of patients with amnesic mild cognitive impairment (MCI) progress to AD per year [2]. However, in the early stages of the disease, the chance of slowing its progression is significantly higher. Several medications are approved by the US Food and Drug Administration (FDA) to delay the onset of symptoms and slow down the progression of dementia at its earliest prodromal phase; however, there is no absolute treatment for AD at this time. Thus the potential for early curative or therapeutic intervention provides added credence and significance to the accurate diagnosis of the early mild cognitive impairment (EMCI) as an early stage in the prognosis of AD.

Diagnosis of AD is mostly based on the clinical history and some neuropsychological tests such as Mini-mental state examination (MMSE) and clinical dementia rating (CDR). However, to understand brain pathology, modern imaging techniques, such as magnetic resonance imaging (MRI) [3, 4], positron emission tomography (PET) [5–8], and functional magnetic resonance imaging (fMRI) [9–12] used as complementary imaging modalities to explore the functional and metabolic interactions among the different brain regions. Moreover, researchers use other biomarkers such as apolipoprotein (APOE) genotype [13, 14], age, sex, and clinical and cognitive test scores [15, 16]. These neuroimaging techniques are non-invasive and provide valuable information for both clinical and research purposes. In addition, multimodal neuroimaging uses the combination of biomarkers to classify and predict AD and its different conversion phases [17–21] which is a great platform for consolidating different measures that allows us to understand the causality of AD and its different prodromal stages.

However, the high dimensionality of these neuroimaging datasets compounded with the low number of multimodal neuroimaging samples available makes the analysis of these types of data quite challenging. Patterns of neuronal cell death, at least in the early stages of the disease, may not necessarily reflect the anatomical or functional abnormalities in the different regions of the brain. Therefore, the analysis should not only carefully scrutinize the different brain regions, with added focus on regions that are known to be disease prone but also look at all potential biomarkers that are best suited to be combined in an optimal fashion to detect these subtle changes. In order to overcome such a problem, machine learning techniques were introduced to analyze medical images data [17, 22–25]. Machine learning algorithms try to find a low dimensional representation of the data which is embedded in high dimensional space. These algorithms using discriminative features

as MCI or AD biomarkers provide powerful models for computer-aided diagnosis for Alzheimer's [26, 27].

Recently, several machine learning techniques have been proposed for the detection of AD and its prodromal stages; among them which are viewed as the most accurate and most applicable approaches are artificial neural network (ANN), support vector machine (SVM), and Bayesian network (BN). SVM is one of the most popular supervised machine learning models with associated learning algorithms to analyze the data applied for classification [28–32]. Bayesian prediction and classification models are another type of machine learning based on the Bayes theorem with the assumption of strong independence among the classification features [33–36]. Deep neural networks and deep learning approach are able to analyze the high dimensional data such as MRI which has been widely used in recent decades [17, 37–43].

SVM remains the most popular supervised machine learning model with associated learning algorithms to analyze the data applied for classification. Through a set of training examples, the SVM algorithm makes a model that is capable of assigning new test data to one of the predefined classes. Cui et al. used anatomical atrophy features extracted from MRI and anisotropy values extracted from diffusion images achieved an accuracy of 71.09% for CN vs. MCI classification by applying an SVM-based method [44]. Suk et al. introduced a deep learning multi-kernel SVM for classification of AD and MCI by combining MRI and PET imaging [40]. In their approach, a set of autoencoders was trained for each imaging modality; then, the obtained high-level features were applied to the multi-kernel SVM for the classification. They obtained 90.7% accuracy for CN vs. MCI and 98.8% for CN vs. AD classification using semi-supervised learning. Furthermore, different techniques based on DNN have been proposed to detect AD with a focus

placed on the MCI group. Liu et al. introduced a new DNN using PET and MRI data and have reported the accuracy of 82.1% for MCI detection from CN group [17]. Ortiz et al. on the other hand introduced a deep belief neural network binary classification and achieved 83% accuracy for the CN vs. MCI converter groups and 90% accuracy for the CN vs. AD groups [45].

Bayesian prediction and classification models are another type of machine learning based on the Bayes theorem with the assumption of strong independence among the classification features. Using such prediction methods, Plant et al. achieved 75% prediction accuracy of the MCI to AD conversion using a combination of voting feature intervals, SVM, and Bayesian model statistics [36]. Through this combination of the Bayesian-based model and the feature selection process, they obtained an accuracy of 92% in the classification of AD and CN groups and 85.71% in the classification of CN and MCI groups. Some other studies used Gaussian analysis which is based on Bayesian classification to perform classification of MCI and AD [46, 47]. Fang et al. introduced Gaussian discriminative analysis (GDA) for early detection of AD and achieved 87.43% accuracy of MCI and AD, 94.1% accuracy of CN and MCI, and 96.92% accuracy of CN and AD classifications using MRI data [46]. In another study by Challis et al., the Bayesian Gaussian process-linear regression model is applied on fMRI data and they report a classification accuracy of 75% for the detection of amnesic MCI from CN and a 97% classification accuracy of aMCI from AD group [47].

Several investigations are reported on the specific diagnosis of MCI, and high classification accuracy of the CN control group from the MCI group. Although such approaches are commendable and have their own merit in shedding light on the progression of the disease, the research community understands that for a disease like AD, the disease may

have started for over a decade prior to any noticeable physical symptoms [48–51]. The complexity of this challenge in delineating the MCI group from the CN group is reflected in the type of classification results, often not exceeding the lower end of the 80% range, that several studies have endeavored to resolve. These attempts, regardless of the multimodal imaging approach and the integration of the different biomarkers along with the use of neuropsychological test scores not initially used at baseline could not discover any new measures that could potentially increase these classification results.

It is thus imperative to include the EMCI group in any prediction or classification study that is bound to assess the different progression phases of the disease with the intent to diagnose the disease in the earliest stage possible. This demands careful examination all types of measures, structural, functional or metabolic, neuropsychological, demographic, and genetic to assess which measures characterize the best progression from normal control to this early stage of mild cognitive impairment. There are current research efforts that attempt at establishing new neuropsychological tests and new imaging techniques that could even precede the EMCI phase [52]. Since EMCI can be considered as an earlier state of mild cognitive impairment, it is of great significance to detect this state for potential early treatment planning and for designing subject-specific early curative/therapeutic intervention protocols. The EMCI stage has shown a milder degree of cognitive impairment as compared to the MCI group, making this phase of the disease more amenable to treatment or to a potential preventive process that will decelerate its progression and provide a longer and better quality of life for these patients; recall the aforementioned percentage of MCI patients that decline to AD.

So far, only a limited number of studies have considered EMCI and LMCI groups [8, 53–59]. Prasad et al. proposed an SVM model to rank brain connectivity features

based on their importance in the classification process [53]. Using diffusion-weighted MRI together with connectivity metrics, an accuracy of 78.2% for CN vs. AD and of 59.2% for CN vs. EMCI classification were obtained by applying an SVM-based classification. They focused more on exploring features that are predictive of AD and used the classification process to better assess the information attained through the connectivity maps. Guerrero et al. reported a higher 65% accuracy for CN vs. EMCI classification using data from the ADNI-GO dataset and making use of the sparse regression for variable selection and manifold learning as a classifier [56]. They used mini-mental state examination (MMSE) instead of disease labels to have a more continuous correlation of the disease stage and SVM with the linear kernel as the classification model.

Singh et al. proposed a feedforward deep neural network to perform classification on fluorodeoxyglucose positron emission tomography (FDG-PET) [8]. They used probabilistic principal component analysis (PPCA) on max-pooled data from FDG-PET and some demographic information including age, gender, APOE ϵ 1, and ϵ 2 alleles, and functional activity questionnaire (FAQ). They achieved a maximum F1-score of 72% for the CN vs. EMCI classification and a 98.14% accuracy for the CN vs. AD classification. Goryawala et al. introduced a linear discriminative analysis (LDA) classifier with two-fold cross-validation using MRI data, demographic information, and neuropsychological test scores [54]. Using MRI features they achieved an accuracy of 61.6% for CN vs. EMCI and 84.2% for CN vs. AD classification. Moreover [60] used MRI features combined with MMSE to determine that the two most discriminative volumetric variables were the right hippocampus and the left inferior lateral ventricle and when combined with MMSE scores provided an average accuracy of 92.4% (sensitivity: 84.0%; specificity: 96.1%) for AD vs. CN classification. Their results also show for amnesic MCI (aMCI) and non-amnesic MCI (naMCI) that brain atrophy is almost evenly seen on both sides of the

brain for AD subjects, which is different from right-side dominance for aMCI and left-side dominance for naMCI. However, since the ADNI subject's diagnosis is based on some neuropsychological tests such as MMSE, involving this parameter trains the algorithm based on clinical ground truth, which evidently increased the accuracy. Shakeri et al. obtained an accuracy of 56% for the CN vs. EMCI classification on MRI data using a multilayer perceptron (MLP) on top of a so-called deep variational autoencoder (VAE) for feature selection and classification [59]. Guo et al. proposed an approach using functional connectivity networks among different brain regions using fMRI data and a multi-kernel SVM classifier that combines multiple variations of functional MRI (fMRI) data [58]. This approach resulted in an accuracy of 72.8% for the classification of the CN vs. EMCI groups and 88.9% for the classification of the CN vs. AD groups; however, one drawback of this study is that their results were based on a limited number of CN, EMCI, and AD subjects (28 CN, 29 AD, and 33 EMCI). Jie et al. used multi-kernel SVM with t-test feature selection algorithm for classification of fMRI data and obtained 66.0% accuracy with 71.4% sensitivity for delineating the CN vs. EMCI [61].

This research endeavor develops the architecture for three machine learning platforms that are based on (1) deep neural network (DNN), (2) support vector machine (SVM), and (3) Gaussian process-based (GP) model classifiers all optimized in order to delineate the different stages of the disease. Radial basis function (RBF) kernel is exploited for classifying four different groups of CN, EMCI, LMCI, and AD. For classification purposes, a DNN approach is proposed for both binary and multiclass classification of CN, EMCI, LMCI, and AD using multimodal neuroimaging (MRI and PET), and other measures that include neuropsychological test scores and demographic data that includes age and education level. In the developed algorithms, we used the Adam optimization technique to update the DNN learning weights. To the best of our knowledge, there are but a handful

of studies that report multiclass classification involving both EMCI and LMCI groups included with the CN and AD groups. Our focus is on the delineation of the EMCI group from the CN group due to the aforementioned importance of early detection, while most studies focused on the MCI diagnosis, combining both EMCI and LMCI groups. We propose a feature selection based on ranking the most important features to help clinicians determine the most essential features in classifying the EMCI group using a large number of subjects in the 4 groups (CN, EMCI, LMCI, AD) considered.

In addition to cross-sectional analysis for the classification of different groups, we introduce a regression framework to predict the future cognitive status of individuals by predicting their cognitive test scores using longitudinal data. Some neuropsychological tests have been designed to assess the clinical status of patients. The Mini-Mental State Examination (MMSE), Clinical Dementia Rating (CDR), Alzheimer’s Disease Assessment Scale Cognitive Total Score (ADAS-cog), and Rey’s Auditory Verbal Learning (RAVLT) are surrogate measures that can be predicted using neuroimaging in order to gauge cognitive decline [62,63]. We present a regression analysis on a 4-year longitudinal study as means to predict future cognitive scores in the trajectory of the disease. This type of regression analysis could play a significant role in pre- and post-treatment evaluations by assessing how disease trajectories are affected to see whether the decline is slowed by such treatment or therapeutic intervention. The focus is placed on the early prediction of conversion from the CN group to MCI with comparatively high accuracy. The challenges for these two groups (CN, EMCI) are the subtle differences that exist between them not only in terms of structural (MRI) functional or metabolic (PET) but also in terms of the small differences that exist in their cognitive scores.

In attempt to predict the brain structural and functional alternations across AD progression, several models have been proposed by researchers. Biomarkers Cascade Model proposed by Jack et al. to predict the abnormalities of FDG- PET at least 24 months before the onset of AD [64] and investigate the tau PET accumulation of the brain in cognitively normal and AD groups [65]. A probabilistic approach based on random vector field transformations has been applied on MRI in order to detect and visualize structural abnormalities in AD and CN groups [66, 67]. A linear regression method was proposed to predict the progression for one year from the MR image at baseline [68]. However, the prediction of progression of Alzheimer’s especially in early stage or presymptomatic remains a challenge [69, 70]. One of the main challenges of early prediction of Alzheimer’s is in the ability to decipher the variability and interrelatedness of the multimodal measurements. Moreover, there are two types of inter-subject variability related to brain changes of the same subject in time and intra-subject variability associated with differences between subjects for the same variable. In addition, time spacing of data acquisition varies even for the same subjects, and age of participants is not temporally aligned. Thus, the unbalanced longitudinal data analysis, especially in multimodal approach could be most challenging [6, 68, 71–74]. To overcome such challenges, a variety of approaches have been proposed such as considering the age of participants [6], using cognitive scores [68, 74], and creating a disease progression score [73, 75–78].

One challenge of longitudinal studies using neuroimaging data is the high dimensionality, which could lead to heavy computational requirements and an ineffective feature selection process. Therefore, sparsity-inducing approaches have been proposed to better identify the predictors by initially reducing the high dimensionality of the problem at hand. Recently multitask learning with sparsity-inducing techniques have been extensively investigated through the shared representation of different tasks with the intent to

minimize the prediction error and identify correctly the relevant features [79–81]. Most of the multitask approaches assume that different tasks are related and share a common set of features [82,83] or share a common subspace [84,85], or the tasks are clustered into groups [86,87]. Wang et al. used l_1 -norm regularizer to constrain the sparsity and applied a combination of l_1 -norm and $l_{2,1}$ -norm regularization to select a subset of features [80]. In another study, they used multitask learning at different time points of a 24-month time period and utilized the lasso regularization in order to predict the MMSE and ADAS scores [88]. A temporally constrained group sparse learning was proposed by Jie et al. to predict ADAS and MMSE on longitudinal data for 24 months [89]. Liu et al. proposed a multi-task feature learning based on sparsity-inducing $l_{2,1}$ -norm to predict eleven neuropsychological tests including MMSE based on MRI [90]. Cao et al. used $l_{2,1}$ -norm penalty applied on regression weights for feature selection and a group $l_{2,1}$ -norm penalty on MRI features to find the cognition-relevant brain regions in a shared subspace [91]. Later in 2018, they proposed a $l_{2,1}$, l_1 -norm to improve the accuracy [92].

Although these single modality studies have merit in assessing the structural or functional alternations in the brain, they reveal that classification and prediction accuracy could be enhanced by consolidating the strength each modality brings in a multimodal processing approach [43, 91, 93–103]. A multimodal multitask learning with $l_{2,1}$ -norm was proposed by Zhang et al. which considered a common feature subset for related tasks using a support vector regression to fuse the features of different modalities [79]. Xiang et al. proposed a multisource sparse regression method focusing on missing modality issue on cross-sectional data [104]. Zhou et al. proposed a multitask learning process which considers the prediction of ADAS and MMSE at each time point in a task using MRI, demographic information, and APOE genotype [105]. They used temporal group lasso regularizer which consisted of two temporal smoothness processes and $l_{2,1}$ -norm penalty

term to ensure successive small deviation between regression models and selection of small subset of variables at each time point, respectively. However, the limitation of this study is that convex fused sparse group lasso (cFSGL) is considering two successive tasks or time points which may lead to losing task dependency between tasks or time points.

There are some studies that use prior knowledge regression models for intra-group similarity using the group information [106, 107]. Nie et al. proposed an unsupervised multisource multitask learning technique to learn MMSE and ADAS simultaneously by using temporal smoothness and prior knowledge of source consistency [108]. However, these methods have some limitations as the learning process for all tasks is done simultaneously for a common subset of brain regions while each task could be trained more optimally using different brain regions. On the other hand, a common limitation that many studies face is in assuming a linear relationship between predicted neuropsychological tests and the extracted neuroimaging measures. In addition, the majority of recent studies were not able to address the missing data challenge, resorting instead to filling for the missing values by interpolation techniques or removing subjects with missing data; a process which ends up limiting the statistical meaningfulness of the study. Although recent studies have considered MCI as the transitional stage from cognitively normal to Alzheimer's disease, this group remains the target of prediction in the progression to AD. However, any treatment or therapy trials are likely to be more successful in the earliest stage of the disease or even in the presymptomatic stage where no manifestation of cognitive decline is yet apparent [48]. Therefore, identification of cognitively normal individuals who are likely to develop into MCI and potentially to AD could increase the chances for intervening early with treatment or therapy and slow the progression of AD.

In order to overcome the aforementioned challenges, a multimodal multitask learning framework is proposed in chapter 5 by exploiting the graph Laplacian regularization as it pertains to sparsity and by considering the task dependency matrix based on Gaussian kernel across all tasks for modality-specific regression. This proposed framework uses a kernel-based regression to investigate the relationship between neuroimaging features and the neuropsychological tests as a nonlinear function, while simultaneously handling the missing data challenge using the decision-tree based fusion. Furthermore, conversion from both groups of CN and MCI are investigated the prediction outcomes of cognitive test scores of MMSE, ADAS, and RAVLT in terms of root mean square error (RMSE) and correlation coefficient (R^2).

CHAPTER 2

A DEEP NEURAL NETWORK APPROACH FOR EARLY DETECTION OF MILD COGNITIVE IMPAIRMENT USING MULTIPLE FEATURES

2.1 Goal

This chapter proposes a machine learning approach based on a deep neural network (DNN) in order to detect AD in its early stage using multimodal imaging, including magnetic resonance imaging (MRI), positron emission tomography (PET) and standard neuropsychological test scores. The proposed approach makes use of the optimization method of Adam to update the learning weights to improve its accuracy. The algorithm is able to classify the cognitively normal control group from early mild cognitive impairment (EMCI) with an unprecedented accuracy of 84.0%. Although the focus here is distinguishing the two groups of CN and EMCI for early diagnosis and treatment planning, this study also shows how the proposed deep learning algorithm can be extended for multiclass classification involving CN and all the stages of EMCI, late MCI (LMCI) and AD. Our approach is able to diagnose EMCI with comparatively high accuracy both for binary and multiclass modes. Based on our knowledge, this is the first time that multiclass involving both EMCI and LMCI classification is reported. In addition, data pre-processing is performed by extracting features from MRI and extracting standard uptake value ration (SUVR) features from registered PET imaging.

2.2 Materials and Methods

2.2.1 Data Acquisition

Data used in the preparation of this study are obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).¹ The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Identification of biological markers at the early stage of AD will help researchers and clinicians to plan for early treatment and therapeutic interventions. The EMCI subject inclusion criteria are as follows: MMSE scores between 24-30, CDR of 0.5, objective memory loss of 0.5-1.5 SD (standard deviation) below normal according to the education adjusted scores on delayed recall of one paragraph from Wechsler Memory Scale Logical Memory II, memory complaints, absence of the significant level of impairment in other cognitive domain, absence of dementia, and essentially preserved activities of daily living. LMCI criteria are almost similar to the EMCI except for the memory loss scores by Wechsler Memory Scale Logical Memory II which is set at more than 1.5 SD below normal.

In this study, a total of 896 participants were classified into the four categories of CN (248), AD (159), EMCI (296), and late mild cognitive impairment (LMCI) (193). All

¹Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

subjects had an MRI and a Florbetapir (18F-AV-45) amyloid PET scan within a 6-month time window. Table 2.1 shows the details of the demographic information of the subjects used in this study. In this research endeavor, we combined the neuroimaging data with some neuropsychological test scores of the subjects, which include Rey’s auditory verbal learning test (RAVLT) as a well-known test for episodic memory, Montreal cognitive assessment (MoCA) known to be effective in the setting of mild cognitive impairment, and everyday cognition total (ECogT) which is associated with measures of global cognition and functional status. Table 2.2 provides the neuropsychological test information of the participants that are used in this study.

Table 2.1: Demographic information of the participants

	CN	EMCI	LMCI	AD
Number of subjects	248	296	193	159
Female/Male	125/123	132/164	84/109	67/92
Age-PET(year)*	75.7(6.5)	71.5(7.4)	73.8(8.1)	74.9(7.8)
Age-MRI (year)	75.2(6.5)	71.3(7.4)	73.6(8.1)	74.7(7.8)
Years of Education	16.4(2.5)	15.9(2.6)	16.2(2.7)	15.7(2.7)

*Values illustrate mean (standard deviation) for all attributes

Table 2.2: Neuropsychological information of the Participants (PF: Percent-Forgetting, Im: Immediate)

	CN	EMCI	LMCI	AD
MoCA	25.7(2.5)	24(2.8)	22.4(3.3)	17.2(4.5)
ECogT	1.2(0.3)	1.6(0.5)	1.9(0.7)	2.8(0.6)
RAVLT-PF	36.2(20.5)	47(30)	67.2(31.5)	89.4(20.5)
RAVLT-Im	45.3(10.5)	40(10.6)	33.2(11)	22.7(7)

2.2.2 MRI Processing

MRI images are 3T T1-weighted using a 3D sagittal volumetric magnetization prepared rapid gradient echo (MP-RAGE) sequence. Here, the repetition time is 2,300 *ms*,

minimum full echo time, inversion time is 900 *ms* with a $256 \times 256 \times 170$ acquisition matrix providing a voxel size of $1.0 \times 1.0 \times 1.2 \text{ mm}^3$. In this study, the T1-weighted standard MNI 305 space MRI has been used to do feature extraction using FreeSurfer version 5.3. The T1-weighted image was used as the reference image in the registration process, which included skull-stripping, segmentation, and demarcation of the different brain regions.

We extracted 3 measurements that included mean intensity, volume, and intensity standard deviation after dividing the image into 45 subcortical regions. In addition, 9 morphological variables including gray matter volume, rectified mean curvature, folding index, surface area, intrinsic curvature index, average thickness, rectified Gaussian curvature index, white matter volume, and thickness standard deviation for 68 cortical regions were generated. Furthermore, the estimated total intracranial volume (eTIV) is calculated, which is used later for normalization of the volumetric measures.

2.2.3 AV-45 PET Processing

PET images used in this study were acquired of 370 MBq (10 mCi), dynamic 3D scan of four 5-minute frames from 50 to 70 minutes post-injection, co-registered, averaged, reoriented into a standard $160 \times 160 \times 96$ voxel image grid with 1.5 *mm* cube voxels, and smoothed to a uniform isotropic resolution of 8 *mm* full width of maximum. In order to acquire the standard uptake value ratio (SUVR) for each participant's amyloid accumulation, first, the AV-45 PET scan was linearly registered onto the T1-weighted image using FSL [109] with 12 degrees of freedom (DoF). The steps of feature extraction from MRI and PET data through image processing are illustrated in the early part of Figure 2.1.

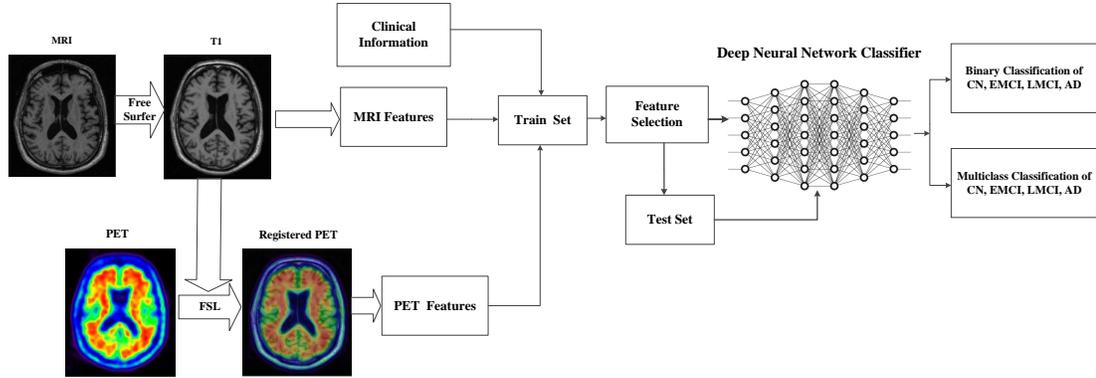


Figure 2.1: Overall diagram of the proposed DNN Classifier.

The registration phase plays an important role to get as much information as possible from PET scan due to the low resolution of this neuroimaging technique. This registration process provides the same MRI parcellation and segmentation for the AV-45 PET image. Then the mean intensity of each FreeSurfer region (ROI) for the 45 subcortical and the 68 cortical regions were assessed, which together identify the standardized uptake values (SUVs). These values extracted from each region were obtained by volume-weighted means as in Equation (2.1).

$$S_n = \frac{\sum_{j=1}^n I_j}{M} \quad (2.1)$$

where S_n is the mean SUV of region n , with M defining the number of voxels found in region n , and I_j is the intensity of voxel j in the AV-45 PET. The whole cerebellum SUV including 4 subcortical regions (right/left cerebellum and right/left white matter)

and global uptake value consisting of 68 cortical ROIs (34 ROIs for each hemisphere) were calculated as follows:

$$S_{CR} = \frac{S_{R_1} \times V_{R_1} + S_{R_2} \times V_{R_2} + \dots + S_{R_k} \times V_{R_k}}{V_{R_1} + V_{R_2} + \dots + V_{R_k}} \quad (2.2)$$

where the S_{CR} represents the SUV of combined 4 cerebellum and 68 cortical regions, S_{R_i} defines the SUV corresponding to region ROI_i and V_{R_i} is a measure of the volume of ROI_i . In the end, all the computed global SUVs were normalized by the whole cerebellum SUV as the reference to compute standard uptake value ratio (SUVR) as expressed in Equation (2.3).

$$SUVR = \frac{Global\ SUV}{Reference\ SUV} \quad (2.3)$$

2.2.4 Statistical Analysis

As mentioned before, one of the challenges of neuroimaging analysis is the high dimensionality of data, specifically in AD and MCI diagnosis with low samples. Therefore, feature selection plays an important role in preprocessing the data. This study is based on 10-fold cross-validation for better reliability of the system. Analysis of variance (ANOVA) was applied to the training data to calculate the variances of groups which is followed by Bonferroni correction to adjust the P-values (threshold of P-value is 0.05 in this study). Then the correlation of each feature is computed and the highly correlated features were removed for dimensionality reduction purpose. The network was thus tested using the test data and the model achieved from the training process.

2.2.5 Feature Normalization

In this study, we used feature-wise scaling in order to normalize the data and increase the accuracy of classification. For this purpose, each feature linearly transformed to have unit standard deviation and zero mean as follow:

$$x_N = \left(\frac{x - \bar{x}}{std} \right) \quad (2.4)$$

where std represents the standard deviation of each feature vector, x represents the feature vector, x_N defines each normalized feature vector, and \bar{x} represents its mean value.

2.2.6 Deep Neural Network

A deep neural network (DNN) can be obtained by increasing the number of hidden layers. In this study, we used 3 hidden layers. In addition, to tackle the overfitting problem, we used the so-called dropout technique [110]. The main idea in the dropout technique is to randomly eliminate (drop out) units (along with their connections) from the DNN during the training process to prevent units from too much co-adapting. This technique helps to significantly reduce the burden of overfitting and yields major advantages in terms of computational requirements and the resulting outcome in comparison with other regularization approaches. The block diagram of the proposed classification algorithm along with feature extraction steps are presented in Figure 2.1.

Here, we used the DNN structure using backpropagation where its learning weights are getting updated with adaptive moment estimation. Adam is an efficient method that

has advantages of both RMSProp and AdaGrad [111]. The weights update rules for the iteration t for the initial values $M_0 = 0$, $R_0 = 0$, $M_{b_0} = 0$, and $R_{b_0} = 0$ are as follows:

$$\begin{aligned}
M_t &= \alpha_1 M_{t-1} + (1 - \alpha_1) \theta_t \\
M_{b_t} &= \alpha_1 M_{b_{t-1}} + (1 - \alpha_1) \theta_{b_t} \\
R_t &= \alpha_2 R_{t-1} + (1 - \alpha_2) \theta_t^2 \\
R_{b_t} &= \alpha_2 R_{b_{t-1}} + (1 - \alpha_2) \theta_{b_t}^2
\end{aligned} \tag{2.5}$$

where M_t and M_{b_t} are the weights and bias for the momentum, α_1 is the hyper-parameter for the momentum, R_t and R_{b_t} are the weights and bias for RMSProp, and α_2 is the hyper-parameter for the RMSProp. Here, α_1 and α_2 are equal to 0.99 and 0.999, respectively. θ_t , θ_{b_t} are the gradients for the weight and bias. In addition, θ_t^2 indicates $\theta_t \odot \theta_t$ using the mini-batch gradient. In the next step, the corrected weights can be computed as follows:

$$\begin{aligned}
\hat{M}_t &= M_t / (1 - \alpha_1^t) \\
\hat{M}_{b_t} &= M_{b_t} / (1 - \alpha_1^t) \\
\hat{R}_t &= R_t / (1 - \alpha_2^t) \\
\hat{R}_{b_t} &= R_{b_t} / (1 - \alpha_2^t)
\end{aligned} \tag{2.6}$$

Finally, the weights and biases are getting updated as follows:

$$\begin{aligned}
v_t &= v_{t-1} \cdot \eta \cdot \frac{\hat{M}_t}{\sqrt{\hat{R}_t + \epsilon}} \\
b_t &= b_{t-1} \cdot \eta \cdot \frac{\hat{M}_{b_t}}{\sqrt{\hat{R}_{b_t} + \epsilon}}
\end{aligned} \tag{2.7}$$

where v_t represents the weights of the DNN, b_t represents the biases, ϵ is a parameter to prevent any division by zero which is considered 10^{-8} and η is learning rate equal to 0.001

in this study. One advantage of Adam is choosing the stepsize very carefully which can be computed as

$$\Delta_t = \frac{\eta \cdot \hat{M}_t}{\sqrt{\hat{R}_t + \epsilon}} \quad (2.8)$$

The stepsize in Adam rule has 2 upper bounds as follows:

$$\left\{ \begin{array}{ll} |\Delta_t| \leq \eta(1 - \alpha_1) \sqrt{1 - \alpha_2} & \text{if } (1 - \alpha_1) > \sqrt{1 - \alpha_2} \\ |\Delta_t| \leq \eta & \text{otherwise} \end{array} \right. \quad (2.9)$$

where Δ_t is the stepsize, α_1 and $\alpha_2 \in (0, 1)$ are exponential decline rates in order to estimate the moment. The activation function, $\sigma(\cdot)$, used is sigmoid defined as

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.10)$$

2.3 Results and Discussion

In this study, a deep neural network classifier was applied to the multimodal MRI and PET imaging with the focus placed on the diagnosis of EMCI from CN subjects. The binary classification results for 6 different groups using multimodal imaging are listed in Table 2.3. It is clear from these results that the combination of demographic information and neuropsychological test scores (NTS) of the participants together with MRI and PET imaging modalities enhances the accuracy of the classification significantly.

Figure 2.2 displays the ROC associated with the participants given in Table 2.3. As it can be seen, a classification accuracy of 84.0% and sensitivity of 83.2% have been achieved for classification of EMCI and cognitively normal subjects using all the available information of participants, while the best accuracy of 68% for CN vs. EMCI using MRI and PET

modalities alone. Moreover, the accuracy of 96.8% and 84.1% for CN vs. AD and CN vs. LMCI were achieved, respectively; which are higher than most of the results obtained from previous studies. In addition, Table 2.4 and Figure 2.3 provides the comparison between the proposed deep neural network and neural network with one hidden layer with the same number of all nodes in DNN. It is obvious that DNN is more successful for the classification of CN vs. EMCI, LMCI, and AD considering the accuracy and AUC of the ROC.

Table 2.3: Accuracy (Acc), Sensitivity (Sen), and Specificity (Spe) of the DNN classifier for different pairs of binary classification. NTS represents the Neuropsychological Test Scores of RAVLT, ECogT, and MoCA. The values are represented as a percent)

	CN vs. EMCI			CN vs. LMCI			CN vs. AD		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
MRI	61.1	66.5	58.7	64.1	53.2	70.2	82.2	75.0	87.3
PET	58.2	66.1	48.6	66.0	54.4	76.0	88.9	85.7	90.3
MRI+PET	68.0	73.1	63.8	71.7	60.1	80.9	89.6	88.9	90.1
MRI+PET+NTS*	84.0	83.2	84.4	84.1	80.4	87.6	96.8	94.1	98.2
	EMCI vs. LMCI			EMCI vs. AD			LMCI vs. AD		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
MRI	62.6	71.3	52.5	77.1	80.2	71.1	61.5	63.6	62.0
PET	58.8	74.2	41.3	78.0	83.1	65.4	64.5	73.4	54.4
MRI+PET	68.2	78.1	57.5	83.2	80.0	86.7	68.4	74.1	68.8
MRI+PET+NTS	69.5	80.6	60.5	90.3	86.7	92.2	80.2	86.8	71.9

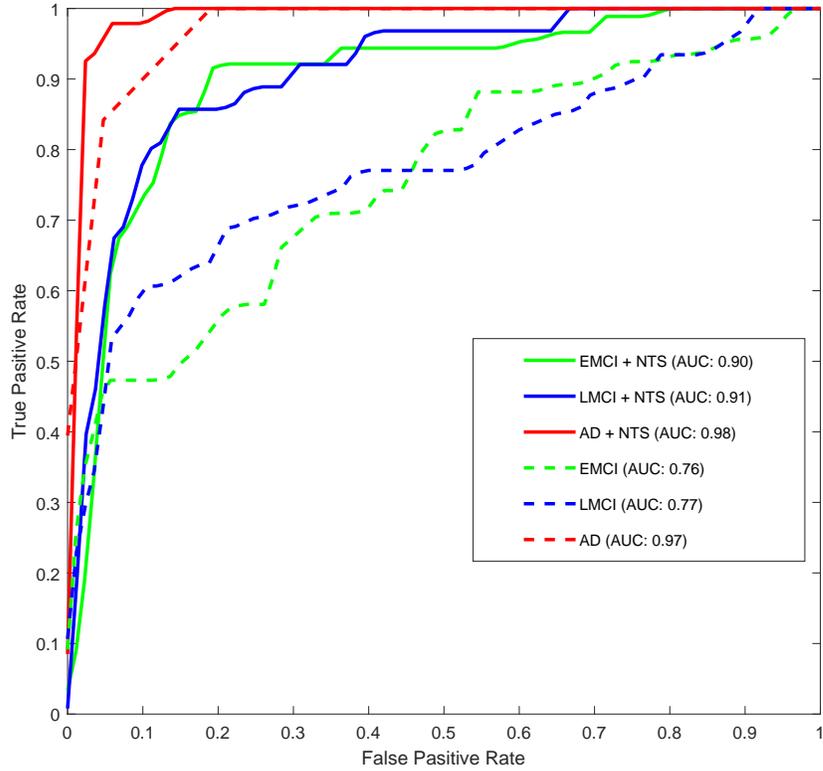


Figure 2.2: ROC curve and area under the curve (AUC) of CN vs. EMCI, LMCI, and AD with and without using neuropsychological test scores for the proposed DNN.

Table 2.4: Accuracy (Acc), Sensitivity (Sen), and Specificity (Spe) of the ANN classifier

	CN vs. EMCI			CN vs. LMCI			CN vs. AD		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
MRI	59.6	74.8	41.9	61.0	53.3	71.7	80.9	80.9	80.2
PET	61.1	69.5	50.9	65.4	56.7	71.7	86.4	81.1	90.3
MRI+PET	64.8	69.3	57.1	67.4	58.6	76.9	89.1	87.4	92.1
MRI+PET+NTS	81.1	84.4	83.1	82.3	78.4	86.5	93.4	91.5	95.1
	EMCI vs. LMCI			EMCI vs. AD			LMCI vs. AD		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
MRI	59.2	69.3	47.0	76.3	70.0	75.1	61.2	67.5	54.2
PET	62.3	78.2	40.0	74.0	66.6	84.1	64.9	68.0	57.4
MRI+PET	61.5	70.1	51.3	79.2	74.2	87.5	66.9	68.8	66.2
MRI+PET+NTS	67.1	79.4	52.1	88.6	83.8	92.5	79.1	81.9	76.1

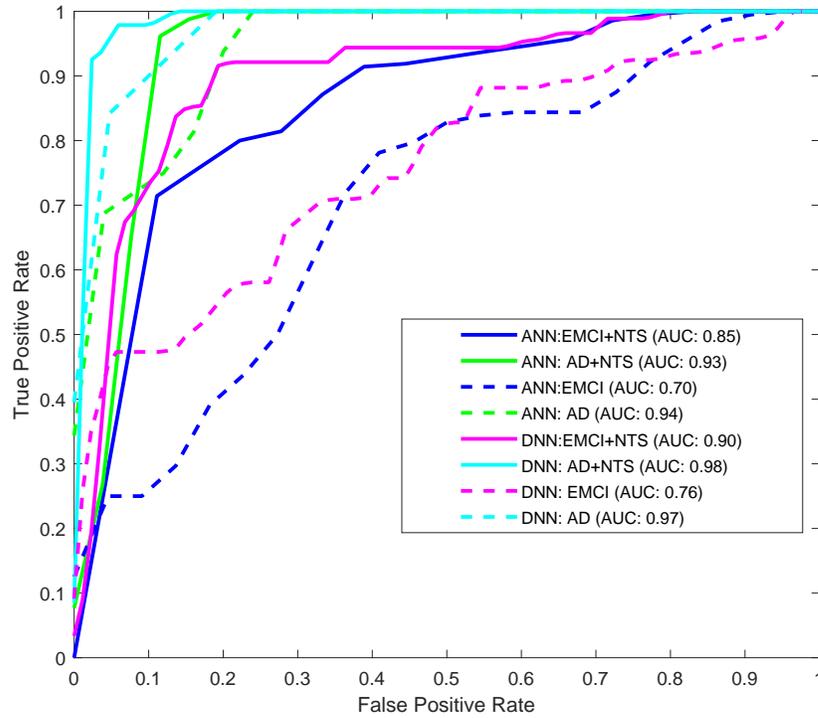


Figure 2.3: ROC curve and AUC values of CN vs. EMCI and CN vs. AD with and without using neuropsychological test scores for both ANN and DNN.

Furthermore, Table 2.5 shows the accuracy resulting from multiclass classification and the classification accuracy for delineating each group from other different groups for ANN and DNN. Although the results for binary classification in some groups are almost similar in ANN and DNN, the accuracy in multiclass classification is higher in the deep neural network; besides, Figure 2.3 indicates that DNN has higher AUC values than ANN, which proves the advantages of DNN over ANN. Here, we used *one vs. all* approach of multiclass classification. Based on our own knowledge, this is the first report of multiclass CN, EMCI, LMCI, and AD classification. Another important point in this study is the high sensitivity in both multiclass and binary classification that were obtained in the EMCI diagnosis. In binary mode, the EMCI was set as the positive class in all pairs.

Table 2.5: Comparing the DNN classifier with ANN considering the overall accuracy (Acc), accuracy of CN detection (Ac1), accuracy of EMCI detection (Ac2), accuracy of LMCI detection (Ac3), and accuracy of AD detection (Ac4) for different groups of one vs. all classification (all the values are in percent and rounded to the nearest integer)

	CN/EMCI/LMCI					CN/EMCI/AD					CN/LMCI/AD				
	Acc	Ac1	Ac2	Ac3	Ac4	Acc	Ac1	Ac2	Ac3	Ac4	Acc	Ac1	Ac2	Ac3	Ac4
ANN	42	39	43	40	-	60	57	55	-	75	55	69	-	33	60
DNN	51	48	57	47	-	61	58	57	-	75	59	72	-	43	59
ANN + NTS	55	69	57	43	-	65	62	58	-	81	64	77	-	48	65
DNN + NTS	60	66	62	50	-	68	66	63	-	78	70	82	-	49	74

	EMCI/LMCI/AD					CN/EMCI/LMCI/AD				
	Acc	Ac1	Ac2	Ac3	Ac4	Acc	Ac1	Ac2	Ac3	Ac4
ANN	51	-	65	30	53	43	47	48	25	48
DNN	54	-	66	35	55	46	52	50	31	53
ANN + NTS	57	-	65	38	69	52	68	50	28	61
DNN + NTS	59	-	73	45	68	57	61	59	39	66

Table 2.6 provided an assessment of our proposed in comparison to related studies, where it can be clearly seen that the proposed method yielded a higher accuracy in almost all the binary classification, and to the best of our knowledge, it is the first study of its kind to report both binary and multiclass classification involving all the groups of CN, EMCI, LMCI, and AD. Goryawala et al. reported the accuracy of 61.6% for CN vs. EMCI classification based on MRI and accuracy of 85.6% for the same groups based on a combination of the MRI data with some neuropsychological tests like MMSE using linear discriminative analysis (LDA) [54]. However, since MMSE is one of the most important criteria for diagnosis in ADNI, involving these parameters means that the algorithm is trained using initially the clinical ground truth. In order to prevent circulation in our method, we tried not to involve the diagnosis parameters of ADNI such as the MMSE score.

Figure 2.4 shows the effects of different tests on three groups of CN vs. EMCI, CN vs. LMCI, and CN vs. AD based on changes in accuracy. We observed that using the RAVLT, ECogT, and MoCA test scores affected the EMCI classification more than any

Table 2.6: Accuracy (Acc) of the DNN classifier comparing to the previous works

	Modality	CN vs. EMCI	CN vs. LMCI	CN vs. AD
Accuracy (%)				
[59]	MRI	56	59	84
[53]	Diffusion MRI	59.2	62.8	78.2
[56]	MRI	65	-	86
[54]	MRI	61.6	71.4	84.2
Proposed	MRI+PET	68.0	71.7	89.6
Proposed	MRI+PET+NTS	84.0	84.1	96.8
	Modality	EMCI vs. LMCI	EMCI vs. AD	LMCI vs. AD
Accuracy (%)				
[59]	MRI	63	81	67
[53]	Diffusion MRI	63.4	-	-
[56]	MRI	-	-	-
[54]	MRI	68.8	81.4	59.6
Proposed	MRI+PET	68.2	83.2	68.4
Proposed	MRI+PET+NTS	69.5	90.3	80.2

other groups based on increasing the accuracy. However, these tests affected the LMCI classification more than any other groups based on increasing the sensitivity as shown in Table 2.3. Using NTS improves the classification accuracy of CN vs. EMCI by almost 16% and CN vs. LMCI by 13% but only by 5% for the CN vs. AD classifications; however, improves the sensitivity of CN vs. LMCI by 20% and CN vs. EMCI by 10%.

It is obvious that among these neuropsychological tests, RAVLT immediate and ECogT contributed more effectively to the EMCI diagnosis. RAVLT scores are capable of reflecting the underlying pathology caused by AD very well and ECogT has contributed to the global cognition and functional status. Therefore, RAVLT and ECogT can be considered as effective early markers for diagnosing the prodromal stage of MCI in people who have memory complaints. In addition, RAVLT percent-forgetting is more contributed to LMCI diagnosis.

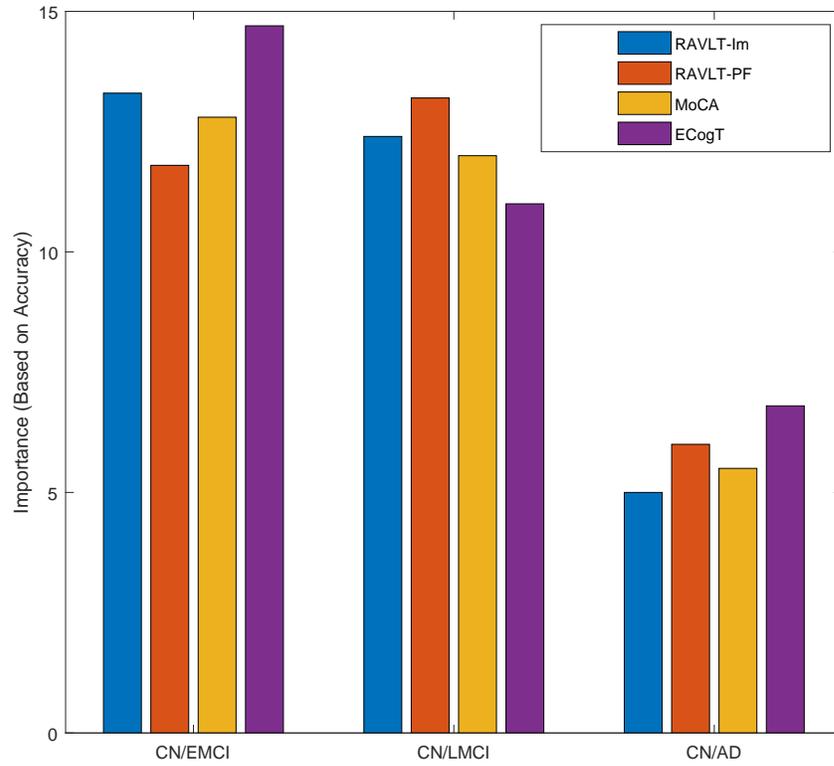


Figure 2.4: Importance of different neuropsychological test scores for three groups of CN vs. EMCI, CN vs. LMCI, and CN vs. AD classifications.

CHAPTER 3

EARLY DIAGNOSIS OF MILD COGNITIVE IMPAIRMENT USING SUPPORT VECTOR MACHINE

3.1 Goal

This chapter proposes a new feature selection algorithm to find the most relevant features. The aim is to use a support vector machine (SVM) approach with radial basis function (RBF) in order to detect AD in its early stage using multiple modalities, including PET, MRI, and standard neuropsychological test scores. A total number of 896 participants were considered in this study. The proposed approach is able to classify cognitively normal control (CN) group from early mild cognitive impairment (EMCI) with an accuracy of 81.1%. In addition, the accuracy of 91.9% for CN vs. late mild cognitive impairment (LMCI) and an accuracy of 96.2% for CN vs. AD classifications were achieved through the proposed model.

3.2 Materials and Methods

3.2.1 Data Acquisition

In this study, we used the same pre-processed data as in Chapter 2. A total number of 896 participants were categorized into four groups of EMCI (296), LMCI (193), CN (248), and AD (159). The overall design architecture of the proposed method is depicted in Figure 3.1.

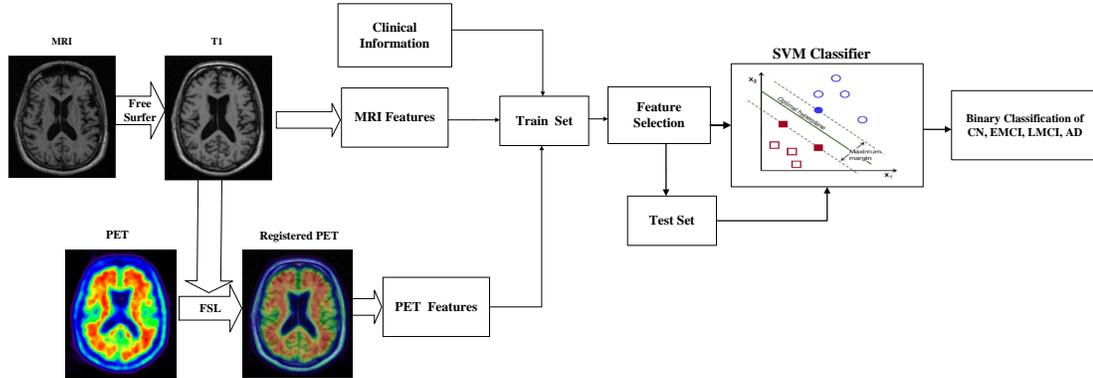


Figure 3.1: Diagram of the preprocessing step and the proposed classification.

3.2.2 Feature Selection

The main challenge in analyzing high-dimensional data is the existence of a very large number of features or variables that may not all be associated with the disease and could be contributing differently at any given phase of the disease. This problem is made even more difficult when the database suffers from a low sample size. Hence, dimensionality reduction techniques or selection of prominent features can play an essential role in machine learning when seeking optimal classification results [112, 113]. In addition, the ranking of these prominent or most relevant features can be appraised whenever the classification results in delineating challenging groups have been optimized, especially when the most subtle of changes differentiate them (like in CN vs. EMCI). These features are deemed relevant only because they do indeed reflect these subtle changes albeit at varying degrees. An optimal decision-making process needs to be established when applying dimension reduction techniques to guarantee that the relevant features are maintained. Although there is a probability to remove some relevant features during feature reduction [114]; however, feature selection techniques are successfully used especially for kernel-based techniques such as SVM [115, 116].

In this study, first, we divided the data into 80% training and 20% testing data and then applied the feature selection process only on training data and assessed the model on the 20% remaining test data. Due to the fact that the random forest (RF) method is time-consuming, we applied ANOVA on the whole training data for each pairwise group separately to reduce the dimensionality considering a P-value of 0.05. Then, we used the RF model to obtain the most important features. Random forest is a tree-based approach, which facilitates multimodal imaging classification by deriving the similarity measures [117]. The RF model combines re-sampling and random feature selection to construct the trees for both classification and regression purposes. On the other hand, RF methodology can provide the hierarchical importance of the different features using statistical permutation and Gini impurity index [118]. The Gini importance score is a measure of variable relevance based on impurity reduction. The Gini impurity, $G(n)$, can be calculated as follows

$$G(j) = 1 - f_{+1}^2 - f_{-1}^2 \quad (3.1)$$

where $f_n = \frac{k_n}{k}$ defines the ratio of the k_n samples from the binary class out of the total samples k at the specific node of j . Then, the reduction of Gini impurity, $\Delta G(j)$ resulting from splitting the samples to j_l and j_r sub-nodes are then calculated as in Equation (3.2)

$$\Delta G(j) = G(j) - f_l G(j_l) - f_r G(j_r) \quad (3.2)$$

where l and r subscripts specify the left and right sub-nodes at each sample splitting. The fraction of data points for the left and right subsets f_l and f_r are defined as $f_l = \frac{k_l}{k}$ and $f_r = \frac{k_r}{k}$. Finally, the Gini Index as an indicator for the selection of feature F is calculated based on aggregating the impurity reduction $\Delta G(j, T)$ for the nodes j and trees T as follow

$$Index(F) = \sum_T \sum_j \Delta G_F(j, T) \quad (3.3)$$

The random forest feature elimination approach is implemented as in the given steps below. First, the algorithm is applied using all variables including age, education, structural and functional variables extracted from the MRI, PET imaging modalities. If we consider p as a sequence of probable number of variables to retain ($p_1 > p_2 > \dots$), at each iteration the variables are ranked based on the explained below process and the top-ranked variables p_j are maintained. The performance of the model is evaluated and the number of variables is determined. Then using the 10-fold cross-validation resampling approach, the above process is encapsulated in one iteration of resampling. This process is repeated for every iteration of 10-fold cross-validation. Then the subsets with the highest accuracy for each iteration were selected and gathered in a pool. In the next step, the features were selected based on the most votes obtained. The RF feature selection model is demonstrated below:

Random Forest Feature Selection

1. for each iteration of 10-fold cross-validation
 - Partition the train data into training and testing sets
 - Train the random forest model using the training set
 - Assess the performance of the model on the testing set
 - Rank the variables according to their importance
 - for each subset of variable numbers p_j for $j = 1, 2, \dots, p$
 - Retain the p_j highest-ranked variables
 - Train the RF model on the training set using p_j variables

- Assess the performance of the model
 - Recalculate the importance of each variable
 - Repeat the process since no variables can be removed
- end
2. end
 3. Choose the subsets with the highest accuracy in each iteration
 4. Gather all selected subsets in a pool and select the variables based on the most votes obtained
 5. Specify the number of variables
 6. Fit the model
-

In the end, we applied correlation versus class labels to the data in order to prevent the algorithm from overfitting by training with too many features. The features with the highest absolute value of correlation are selected based on the Pearson's correlation coefficient.

3.2.3 SVM model

The SVM-based classifier was originally introduced by Vapnik [119] and has been widely used in classification problems due to their robust performance in the presence of noisy data. In SVM, a set of training data defined as: $(y_1, x_1), \dots, (y_k, x_k)$, can be partitioned by the hyperplane of $(w_0, x) + b_0 = 0$, where $w_0 = \sum_{m=1}^k y_m \eta_m^0 x_m$. By substituting w_0 into hyper plane equation, the nonlinear classifier rule can be described as a linear combination of kernels associated with the support vectors as [120]:

$$\sum_{m=1}^k \eta_m^0 y_m K(x, x_m) + b_0 = 0 \quad (3.4)$$

where $y_m \in \{-1, +1\}$ is the corresponding class label, x_m is the training pattern, b_0 is the offset error, k is the number of training samples, and K is the kernel. η is the Lagrange multiplier, and η_m^0 is the solution of the following quadratic optimization problem:

$$W(\eta) = \sum_{m=1}^k \eta_m - \frac{1}{2} \sum_{m,n=1}^k \eta_m \eta_n y_m y_n K(x_m, x_n) \quad (3.5)$$

which is subjected to the following constraints: $\sum_{m=1}^k y_m \eta_m = 0$, $0 \leq \eta_m \leq C$, ($m = 1, \dots, k$), where C is the parameter used to penalize the classification error. In this work, we used Gaussian kernel also known as radial basis function (RBF) [121], that can be computed by Equation (4.4) as follows:

$$[K]_{m,n} = \exp\left(-\frac{\| (x_m - x_n) \|^2}{\sigma^2}\right) \quad (3.6)$$

3.3 Results

In this study, an SVM classifier was applied to the multimodal MRI and PET imaging system with a focus on delineating the EMCI group from the CN group. The results for six different binary classifications using multimodal imaging are listed in Table 3.1. It is clear from these results that the combination of demographic information and neuropsychological test scores (NTS) of the participants together with MRI and PET imaging modalities enhances the accuracy of the classification significantly. As it can be seen from Table 3.1, a classification accuracy of 81.1%, a sensitivity of 82.8% and a specificity of 79.8% have been achieved for classification of EMCI and CN subjects using all the available information of participants. For the classification of subtle changes that exist between such groups (CN, EMCI), these results provide credence to the proposed approach. Moreover, the accuracy of 91.9% and 96.2% for CN vs. LMCI and CN vs. AD were achieved, respectively.

Table 3.1: Accuracy (Acc), Sensitivity (Sen), and Specificity (Spe) of the SVM classifier for different pairs of binary classification (values are represented as percent)

	CN vs. EMCI			CN vs. LMCI			CN vs. AD		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
MRI	73.1	76.8	68.4	63.0	61.5	69.9	90.3	90.3	91.8
PET	-*	-	-	73.6	69.5	83.7	82.5	79.7	83.7
Proposed	75.6	78.9	70.6	76.9	70.0	85.7	91.2	90.1	93.9
Proposed+NTS**	81.1	82.8	79.8	91.9	82.4	97.9	96.2	93.3	97.9
	EMCI vs. LMCI			EMCI vs. AD			LMCI vs. AD		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
MRI	70.1	81.1	55.0	84.7	89.8	80.7	65.2	60.5	70.0
PET	63.9	42.1	78.0	70.8	87.6	55.6	73.2	76.4	61.3
Proposed	70.1	80.0	60.0	85.5	90.5	80.8	78.3	76.3	80.6
Proposed+NTS	71.5	81.3	63.0	93.2	98.0	80.0	80.2	76.3	86.2

* -: No data available

** NTS: Neuropsychological Test Scores of RAVLT, ECogT, and MoCA

Figure 3.2 displays the ROC and area under the curve (AUC) associated with the classification results in Table 3.1. The AUC of 0.81, 0.80, and 0.96 are obtained for classification of CN vs. EMCI, CN vs. LMCI, and CN vs. AD, respectively without involving the cognitive tests as well as AUC of 0.88, 0.98, and 0.99 for classification of CN vs. EMCI, CN vs. LMCI, and CN vs. AD, respectively with the combination of neuroimaging features and cognitive tests.

Figure 3.3 shows the importance of every neuropsychological test scores based on sensitivity changes for three groups of CN vs. EMCI, CN vs. LMCI, and CN vs. AD. We observed that using the NTS including RAVLT, ECogT and MoCA affected the LMCI classification more than any other groups. It is evident from Table 3.1, using NTS improves the accuracy of CN vs. LMCI by almost 15% but only by 6% for the CN vs. EMCI and by 5% for the CN vs. AD classifications.

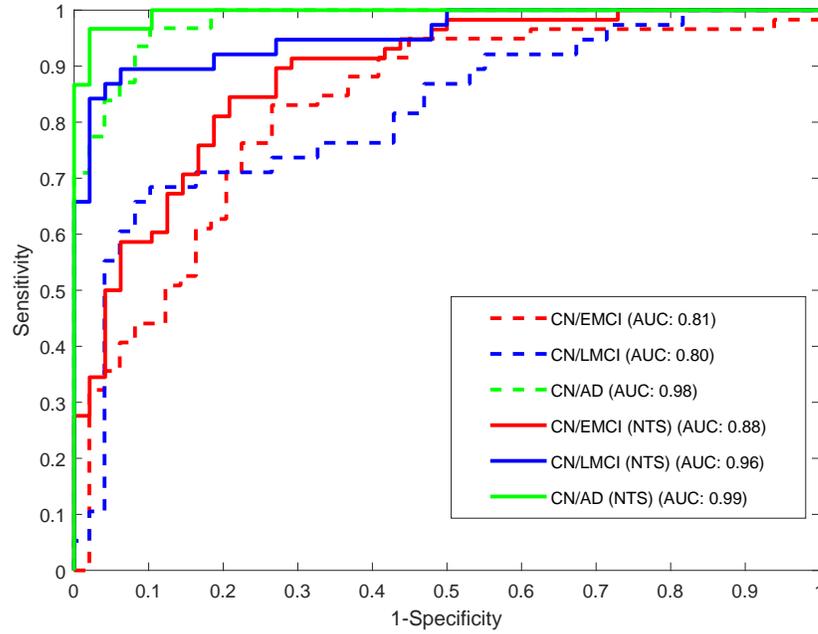


Figure 3.2: ROC curve and AUC of three classifications of CN vs. EMCI, CN vs. LMCI, and CN vs. AD using NTS and without using NTS.

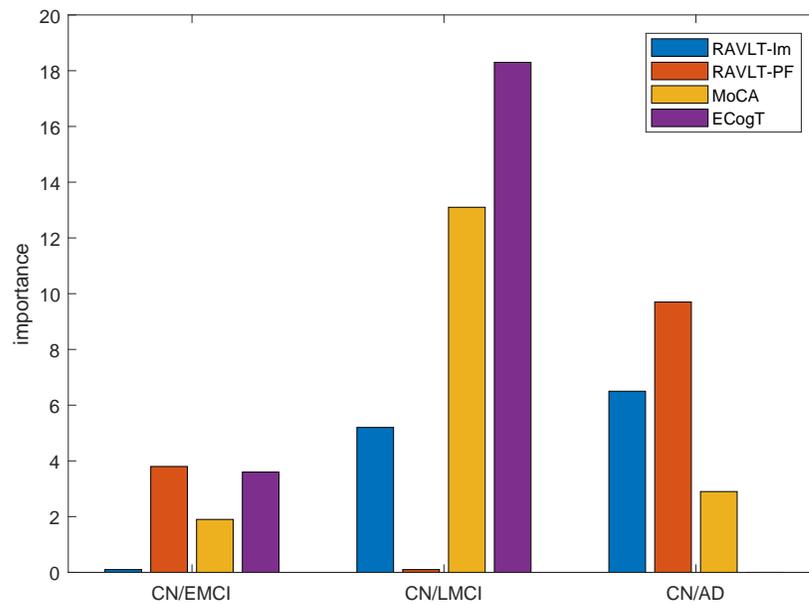


Figure 3.3: Impact of neuropsychological test scores (NTS) used in this study for three classifications of CN vs. EMCI, CN vs. LMCI, and CN vs. AD separately. (PF: Percent-Forgetting, Im: Immediate)

We found that among these neuropsychological tests, ECogT contributed more effectively to the LMCI and RAVLT percent-forgetting to the EMCI diagnosis. MoCA also enhances the classification accuracy of CN vs. LMCI significantly while, interestingly, RAVLT immediate and RAVLT percent-forgetting have no impact on both CN vs. EMCI and CN vs. LMCI classifications, respectively. RAVLT scores are capable of reflecting the underlying pathology caused by AD very well. Therefore, RAVLT percent-forgetting can be considered as an effective marker for the detection of the early stage of Alzheimer’s disease in patients who have memory complaints. In addition, ECogT which is associated with global cognition and functional status is also a useful clinical tool for early and late diagnosis of MCI. It is obvious that CN vs. AD is more affected by RAVLT percent-forgetting in comparison with the other tests. Table 3.2 provides an assessment of our proposed approach in comparison to related studies, where it can be clearly seen that the proposed method with and without neuropsychological test scores yielded a higher accuracy in almost all the binary classification groups.

Table 3.2: Accuracy (Acc) of the proposed method compared to the previous studies of EMCI and LMCI

	CN vs. EMCI	CN vs. LMCI	CN vs. AD
[59]	56	59	84
[53]	59.2	62.8	78.2
[56]	65	-	86
[58]	72.8	78.6	88.9
Proposed	75.6	76.9	91.2
proposed+NTS	81.1	91.9	96.2
	EMCI vs. LMCI	EMCI vs. AD	LMCI vs. AD
[59]	63	81	67
[53]	63.4	-	-
[56]	-	-	-
[58]	-	-	-
Proposed	70.1	85.5	78.3
proposed+NTS	71.5	93.2	80.2

CHAPTER 4

A GAUSSIAN-BASED MODEL FOR EARLY DIAGNOSIS OF MILD COGNITIVE IMPAIRMENT USING MULTIMODAL NEUROIMAGING

4.1 Goal

This chapter develops a random forest feature selection model with a Gaussian-based classifier. This integrated method serves to define multivariate normal distributions in order to classify different stages of AD, with the focus placed on detecting EMCI subjects in the most challenging classification of CN vs. EMCI. Using 896 participants classified into the four categories of CN, EMCI, LMCI, and AD, the results show that the EMCI group can be delineated from the CN group with a relatively high accuracy of 78.8% and sensitivity of 81.3%. Moreover, the performance of the feature selection model and the Gaussian process-based classifier are compared to other state-of-the-art algorithms. The proposed method outperformed others such as minimum redundancy maximum relevance (MRMR) and t-test feature selection methods.

4.2 Materials and Methods

4.2.1 Data Acquisition

In this study, we used the same preprocessed data as in chapter 2. A total number of 896 participants were categorized into four groups of EMCI (296), LMCI (193), CN (248), and AD (159). The overall view of the proposed method is depicted in Figure 4.1. The same feature selection method as chapter 3 was applied to the data in order to assess the performance of the proposed feature selection and classifier with other methods.

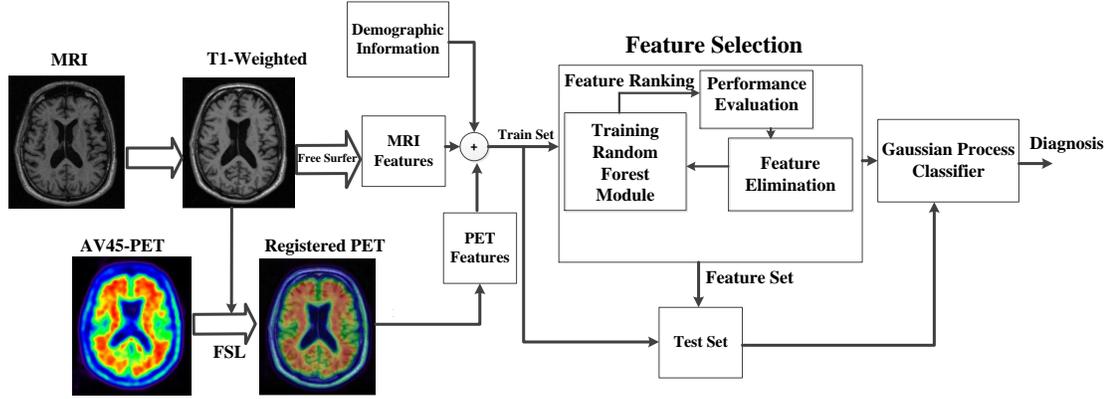


Figure 4.1: Neuroimaging feature extraction and overall diagram of the proposed method.

4.3 Gaussian Process

Gaussian process models (GPs) are a class of supervised machine learning based on Bayes theory for updating probabilities on the assumed hypotheses. Like SVM approaches, GPs are kernel-based, which makes them efficient for high-dimensional data analysis. The Gaussian process as implemented here is a probabilistic approach that uses the average predictive probability instead of a single model. This probabilistic approach can be adapted to the classification problem by transforming the output using the appropriate activation function [122]. The primary goal for training data points x_i for N samples with an associated binary class labeled as $y_i \in \{-1, +1\}$, is to predict the class for which the new testing data points belong. The basic aim of GP classifiers is to predict the probabilities of $p(y|x)$ for the test input. With the following four steps described in Equations (4.3.1) through (4.3.4), we illustrate how to calculate the GP prediction [47].

4.3.1 GP Likelihood

The first step in GP prediction is to define a likelihood for the prediction output. Here, for a binary classification, $y \in \{-1, 1\}$, the probability can be described by a conditional distribution as follows:

$$p(y|z) = (1 + e^{-yz})^{-1} = \sigma(yz) \quad (4.1)$$

where $\sigma(\bullet)$ is the logistic sigmoid function presented in Figure 4.2. As can be seen in this distribution, in each label, the probability can be controlled by the magnitude of the z . The GP considers z as an unobservable variable. The data likelihood can be defined by taking the product over all the classes in the training data as expressed in Equation (4.2)

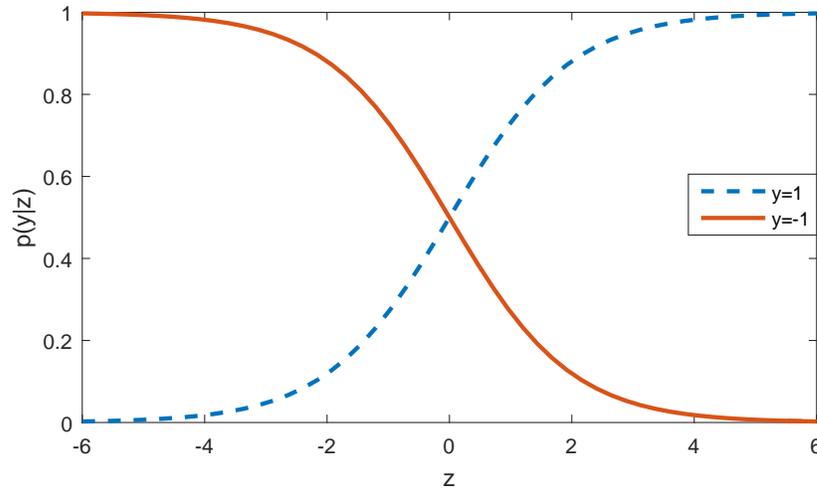


Figure 4.2: The logistic sigmoid function used in the proposed GP learning is plotted based on the latent variable z and class labels of y .

$$p(y|z) = \prod_{i=1}^N p(y^i|z^i) = \prod_{i=1}^N \sigma(y^i z^i) \quad (4.2)$$

where N is the number of training data.

4.3.2 GP Prior

The importance of the specification of GP prior is because of its ability to fix the properties of the functions for the inference. In order to use the specification of GP prior, it is assumed that the unobservable variables, $[z^1, \dots, z^N]^T = z \in R^N$, are distributed based on a GP prior, $g(z|0, K)$, where

$$P(z|X) = g(z|0, K) = \frac{\exp(-\frac{1}{2}z^T K^{-1}z)}{\{\det(2\pi K)\}^{1/2}} \quad (4.3)$$

Here, X is the training input which is the output of feature selection that may contain any of the features of MRI, PET, age, education, with g being the probability density function with zero mean vector and its covariance matrix, $K \in R^{N \times N}$, is symmetric positive-semidefinite with K^{-1} defining the inverse of the covariance matrix. In order to obtain the covariance function of the GP prior, we used Gaussian kernel also known as radial basis function (RBF) that can be computed by Equation (4.4) as follows [121]:

$$[K]_{m,n} = k(x_m, x_n) = \exp\left(-\frac{\|x_m - x_n\|^2}{2\sigma^2}\right) \quad (4.4)$$

where $[K]_{mn}$ denotes the element belongs to row m and column n of the covariance matrix K , k is the covariance kernel, and x_m and x_n are the input vectors that can be represented by RBF kernel as feature vectors.

4.3.3 Marginal likelihood

After obtaining the GP likelihood, $p(y|z)$, and the GP prior, $P(z|X) = g(z|0, K)$, the complete data likelihood can be defined as

$$p(y, z|X) = p(y|z)p(z|X) = g(z|0, K) \prod_{i=1}^N \sigma(y^i z^i) \quad (4.5)$$

Considering the fact that z is latent, and in order to obtain the marginal likelihood, equation 4.5 should be integrated with respect to z as expressed below:

$$p(y|X) = \int p(y, z|X) dz = \int g(z|0, K) \prod_{i=1}^N \sigma(y^i z^i) dz \quad (4.6)$$

The marginal likelihood value can thus be interpreted as the probability of observing the behavior of the training data based on the modeling assumptions.

4.3.4 Posterior distribution

Obtaining the marginal likelihood and the GP prior were the initial steps to obtain the posterior distribution or the predictions for the test set, X' . This distribution is computed using the following equation

$$p(\hat{y}'|X') = \int p(\hat{y}'|z') p(z'|z) p(z|X) dz dz' \quad (4.7)$$

where $p(\hat{y}'|z') = \sigma(\hat{y}'z')$ and the probabilities $p(z|X)$ and $p(z'|z)$ are as defined

$$p(z|X) = \frac{p(y, z|X)}{p(y|X)} \quad (4.8)$$

$$p(z'|z) = g(z'|\sigma' K^{-1} z, \sigma'' - \sigma'^T K^{-1} \sigma') \quad (4.9)$$

Here, $\sigma'' = K(X', X')$ and $\sigma' = K(X', X)$. In order to solve for Equation (4.5) and Equation (4.8), we used Laplace approximation for determining a Gaussian approximation [123]. Having found the posterior distribution, $p(\hat{y}'|X')$, for prediction as well to examine the accuracy of the model. In a binary classification, Equation (4.10) is used to make

a binary prediction. In a binary classification, to make a binary prediction $\hat{y}' \in \{-1, 1\}$ based on the posterior distribution as follows:

$$\hat{y}' = \begin{cases} +1 & \text{if } p(\hat{y}' = 1|X) > \eta, \\ -1 & \text{otherwise} \end{cases} \quad (4.10)$$

where $\eta \in (0, 1)$ is a coefficient used to compensate the class imbalances in the training data which is to equal to 0.5 here. For example, if we wish to make fewer false positive or false negative misclassifications, the threshold parameter, η , can be tuned in between 0 and 1.

4.4 Results

The feature selection process and the Gaussian model were implemented using R software [124] to classify 6 binary groups of [CN vs. EMCI, CN vs. LMCI, CN vs. AD, EMCI vs. LMCI, EMCI vs. AD, and LMCI vs. AD]. Different metrics such as accuracy (Acc), sensitivity (Sen), and specificity (Spe) are determined to assess the performance of the algorithm. Table 4.1 presents the classification results using the different imaging modalities when used separately and when combined. Here, MRI and PET features are selected based on the proposed feature selection algorithm. As can be seen from the results, for the most challenging CN vs. EMCI classification, an accuracy of 78.8%, a sensitivity of 81.4% and a specificity of 76.8% have been obtained when combining MRI, PET, and the demographic information. It should be noted that sensitivity is viewed as the most important metric among these parameters since it reflects the accuracy of diagnosing the true positive group in every binary classification. The EMCI (when considered) is assumed as true positive, $y = +1$, in every pairwise classification. In other pairs, except for CN vs. AD, LMCI is considered as the true positive group.

Table 4.1: Performance comparison of the proposed method for 6 binary groups

Modality	CN vs. EMCI			CN vs. LMCI			CN vs. AD		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
MRI	75.9	77.9	75.5	62.1	48.1	77.5	83.6	80.6	85.7
PET	-*	-	-	76.1	66.4	85.8	90.0	90.3	89.8
MRI+PET	75.9	77.9	75.5	78.1	69.9	87.8	92.5	92.3	93.8
MRI+PET+DI**	78.8	81.3	76.8	79.8	70.2	89.9	94.7	92.3	95.9
Modality	EMCI vs. LMCI			EMCI vs. AD			LMCI vs. AD		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
MRI	72.1	80.3	61.9	85.6	88.8	87.4	62.3	55.3	73.0
PET	62.8	61.5	64.3	69.1	80.3	45.7	69.7	76.0	61.3
MRI+PET	72.5	81.2	66.9	88.1	92.8	87.4	77.1	79.9	75.9
MRI+PET+DI**	73.2	81.2	69.9	88.1	92.8	87.4	77.1	79.9	75.9

*-: No PET features were selected by the RF-RFE process

**DI: Demographic information consists of age and education

In EMCI vs. CN classification, none of the PET features were selected by the proposed RF-RFE process, which may indicate that in this early stage SUVR measurements do not contribute to the classification results. This last assertion indicates that given the low resolution of PET, SUVR measurements are unable to extract the subtle changes that delineate the two groups of CN and EMCI. Except for the CN vs. EMCI classification, combining MRI and PET modalities enhanced all accuracy, sensitivity, and specificity results. Inclusion of age and education level to the multimodal (MRI and PET combination) imaging framework improved the results even further albeit slightly in some cases. Interestingly, from these results, we see that age and education level did not play any significant role in the classification of either EMCI and LMCI groups with AD. Based on our finding in this study, PET features begin to contribute to the overall classification accuracy for the LMCI group. Figure 4.3 shows classification results including Acc, Sen, Spe, the area under the curve (AUC), positive predictive value (PPV), and negative predictive value (NPV) for different modalities separately and combined.

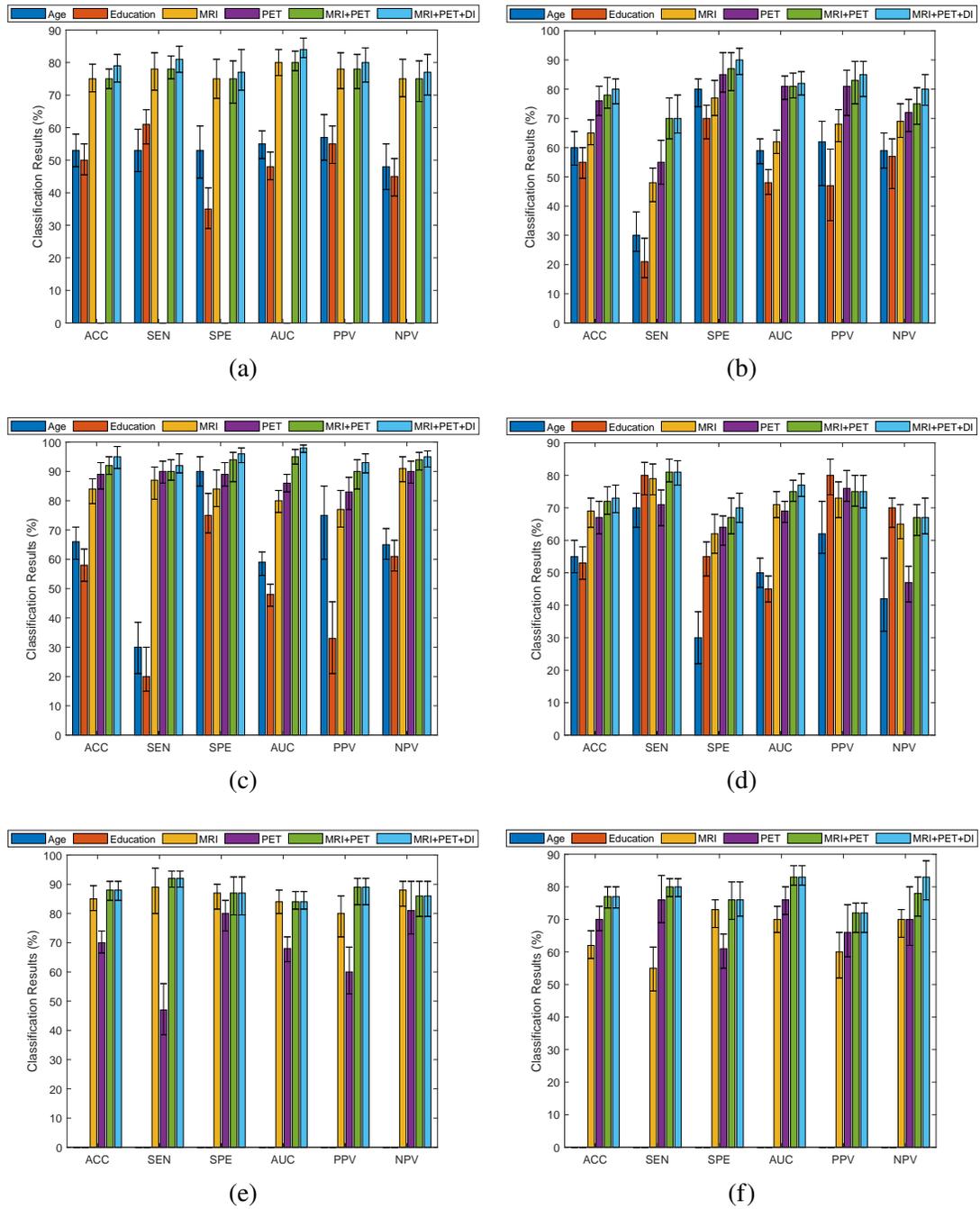


Figure 4.3: Classification results with 95% confidence interval for different modalities for the binary classifications of (a) CN vs. EMCI, (b) CN vs. LMCI, (c) CN vs. AD, (d) EMCI vs. LMCI, (e) EMCI vs. AD, and (f) LMCI vs. AD based on results of Table 4.1.

Figure 4.4 displays the ROC for the pairwise classification of all the considered groups. The area under the plot of ROC (AUC), a plot of true positive rate versus false positive rate, can be a useful tool to evaluate the accuracy of the classifier. An AUC value is between 0 and 1, and an ideal classifier will associate a value of 1. Here, a high AUC of 0.84 was achieved for the challenging CN vs. EMCI classification as mentioned in Figure 4.4. Evidently, and as expected, higher AUC values of 0.98 for the CN vs. AD classification, and of 0.95 for the EMCI vs. AD have been achieved. In addition, AUC of 0.77, 0.82, and 0.83 have been obtained for EMCI vs. LMCI, CN vs. LMCI, and LMCI vs. AD, respectively. An observation that can be made here is that the AUC value for the EMCI vs. LMCI is lower than its counterpart for the CN vs. EMCI classification, which could mean that the neuropsychological test scores used at baseline for this type of delineation (EMCI vs. LMCI) are more relevant than what neuroimaging measures could extract.

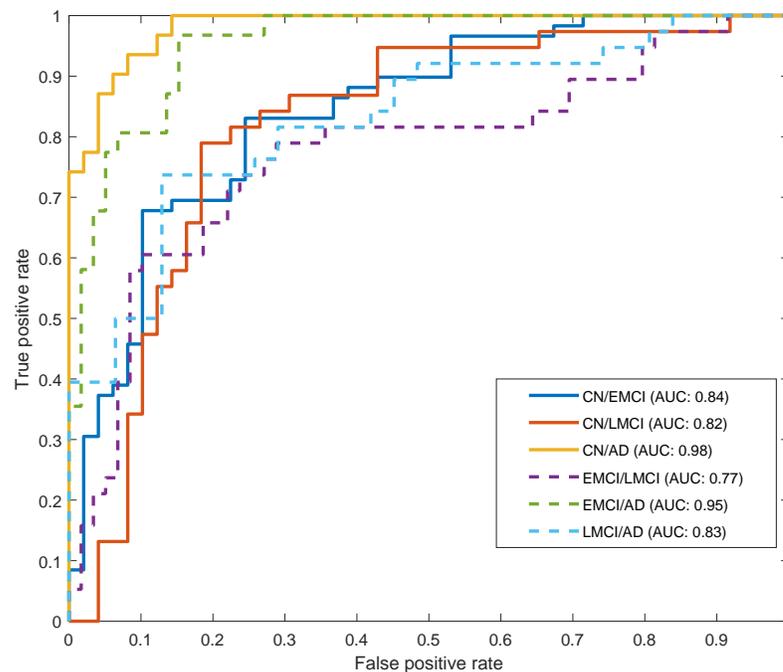


Figure 4.4: Receiver Operating Characteristics (ROC) curve and Area Under the Curve (AUC).

We also investigated what constituted the most important variables which are to be selected for training the classifier. Table 4.2 provides the eight most important features selected by the algorithm along with the P-value related to the analysis of variance. Table 4.2 only provides the eight most important features; however, the number of features that have been selected by the feature selection algorithm is higher for every binary classification group and could be different from the others. The P-values are produced in the first step before applying random forest. It can be observed that different features from MRI and PET data have been selected for the different binary classifications. For example, all features selected, for the CN vs. EMCI classification belonged to the MRI data, while most of the features for the binary classifications of CN vs. LMCI and CN vs. AD were selected from PET data.

Table 4.2: The first eight most important features selected by the feature selection along with the P-value. (lh: left hemisphere, rh: right hemisphere)

CN vs. EMCI	P-Value	CN vs. LMCI	P-Value	CN vs. AD	P-Value
age	$6.06e^{-5}$	lh superior frontal SUVR	$2.03e^{-14}$	lh entorhinal thickness	$1.3e^{-30}$
lh lateral ventricle volume	$3.66e^{-6}$	lh cortical SUVR	$1.47e^{-14}$	rh superior frontal SUVR	$9.49e^{-29}$
lh precuneus volume	$7.24e^{-5}$	lh frontal pole SUVR	$1.76e^{-12}$	lh cortical SUVR	$1.25e^{-29}$
lh superior parietal volume	$7.87e^{-6}$	rh cortical SUVR	$2.73e^{-12}$	rh medial orbitofrontal SUVR	$4.50e^{-27}$
lh superior frontal volume	$1.96e^{-8}$	rh superior frontal SUVR	$1.12e^{-14}$	lh rostral middle frontal SUVR	$6.27e^{-32}$
rh lingual volume	$1.46e^{-6}$	lh parstriangularis SUVR	$1.87e^{-14}$	rh frontal pole SUVR	$4.30e^{-30}$
rh lateral ventricle volume	$8.55e^{-5}$	lh middle temporal SUVR	$1.77e^{-12}$	lh caudal middle frontal SUVR	$1.69e^{-29}$
3rd ventricle volume	$2.31e^{-6}$	lh inferior parietal SUVR	$3.35e^{-11}$	lh accumbens SUVR	$1.44e^{-29}$
EMCI vs. LMCI	P-Value	EMCI vs. AD	P-Value	LMCI vs. AD	P-Value
lh precuneus volume	$1.45e^{-4}$	lh middle temporal volume	$2.50e^{-18}$	rh inferior temporal volume	$4.05e^{-6}$
lh superior frontal SUVR	$9.96e^{-6}$	lh middle temporal SUVR	$4.49e^{-17}$	rh lateral occipital SUVR	$2.70e^{-5}$
lh cortical SUVR	$4.63e^{-6}$	rh middle temporal SUVR	$1.10e^{-20}$	lh amygdala volume	$1.28e^{-7}$
lh hippocampus volume	$3.90e^{-8}$	lh entorhinal thickness	$3.02e^{-25}$	lh pericalcarine SUVR	$1.58e^{-6}$
lh amygdala volume	$1.26e^{-6}$	rh medial orbitofrontal SUVR	$7.92e^{-20}$	lh hippocampus volume	$3.22e^{-7}$
lh frontal pole SUVR	$1.69e^{-4}$	lh hippocampus volume	$1.07e^{-22}$	lh inferior temporal volume	$1.69e^{-5}$
lh parahippocampal volume	$2.82e^{-4}$	rh precauneus volume	$3.16e^{-14}$	rh pericalcarine SUVR	$9.80e^{-6}$

We observed that using multimodal imaging enhances the accuracy differently for each binary classification as a function of the features that were deemed important and from which modality they were extracted from. The results of CN vs. EMCI and CN vs. LMCI suggest that the beta-amyloid deposition in its very early stage of Alzheimer’s disease is not as significant as in the later stage of the disease. The amyloid burden will probably continue to increase during the transition between EMCI and LMCI. Figure 4.5 shows the importance of the features based on the Gini importance measure which are listed in Table 4.2 for the challenging group, CN vs. EMCI. In addition, Figure 4.6 presents box plot of different features from CN vs. EMCI, CN vs. LMCI, and CN vs. AD classifications, indicating the significance of different features at the different stages of the disease.

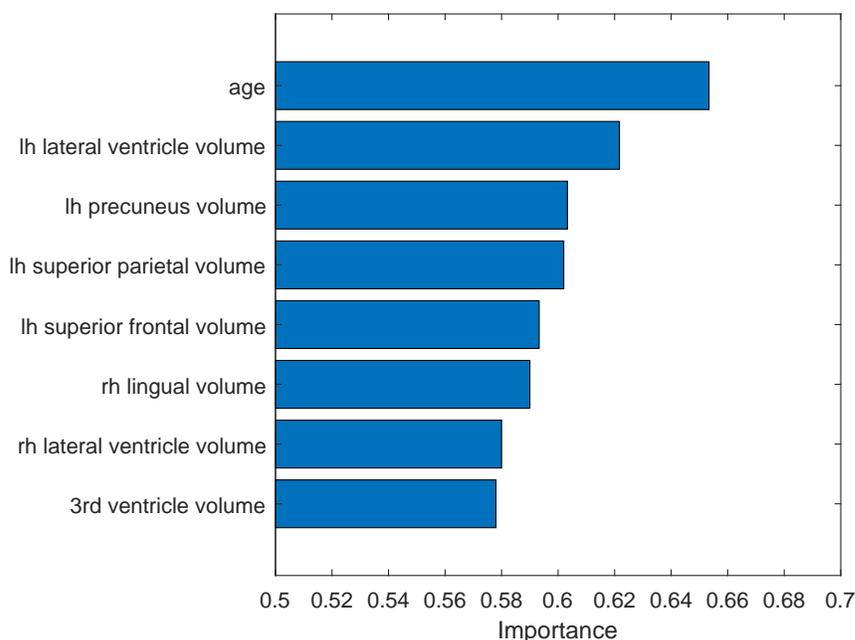
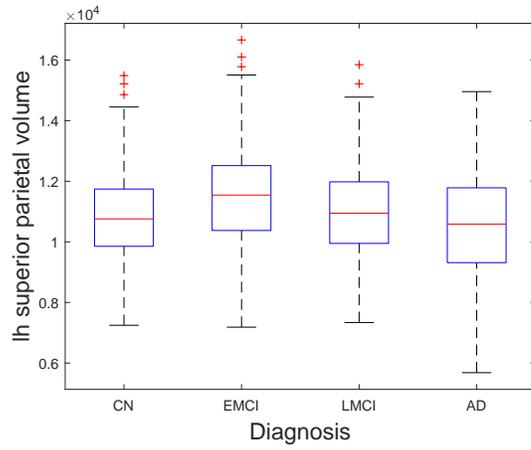
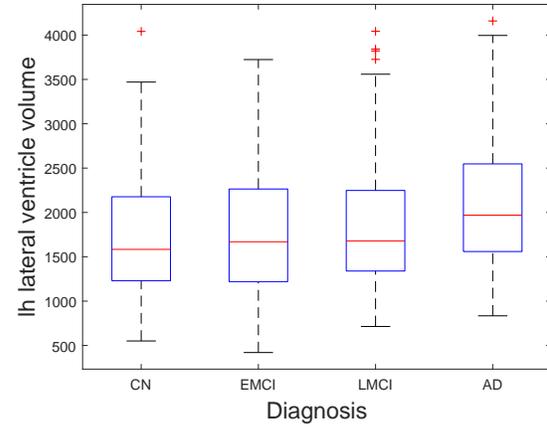


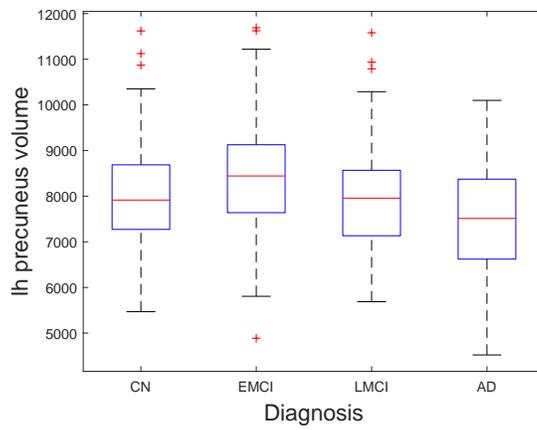
Figure 4.5: Importance of the features listed in Table 4.2 for the most challenging CN vs. EMCI classification.



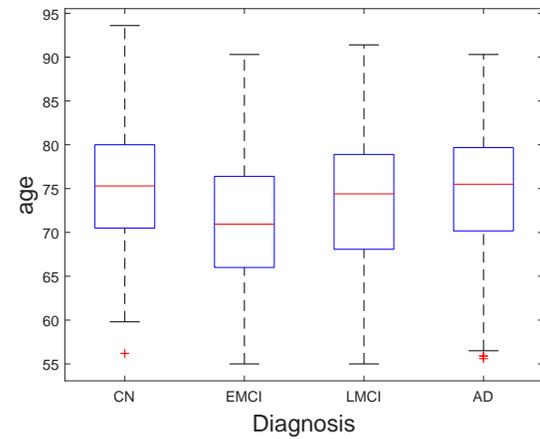
(a)



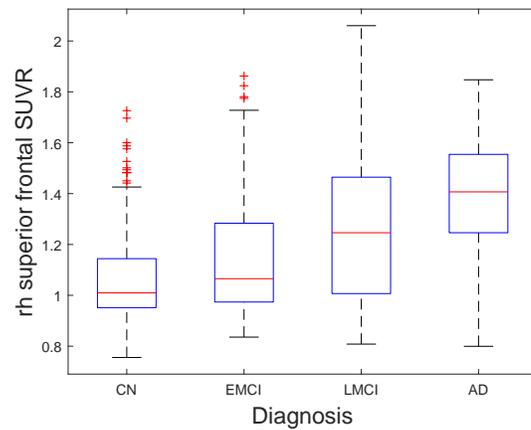
(b)



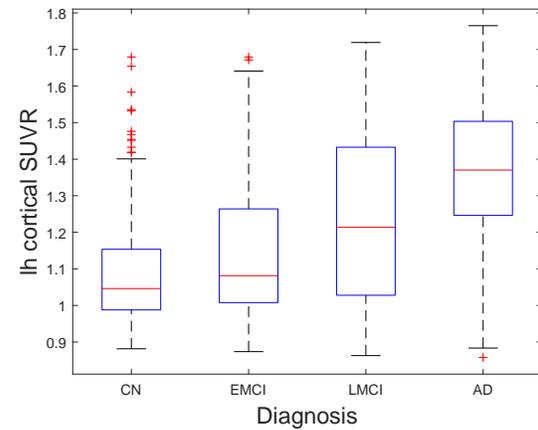
(c)



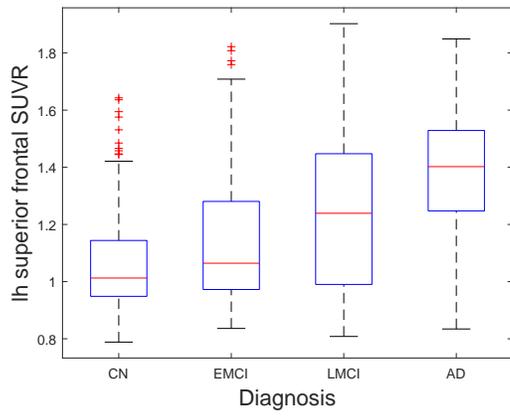
(d)



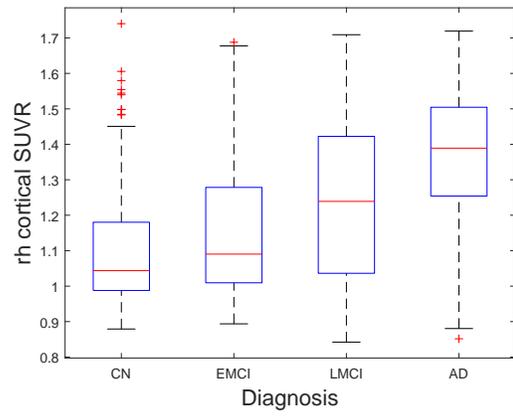
(e)



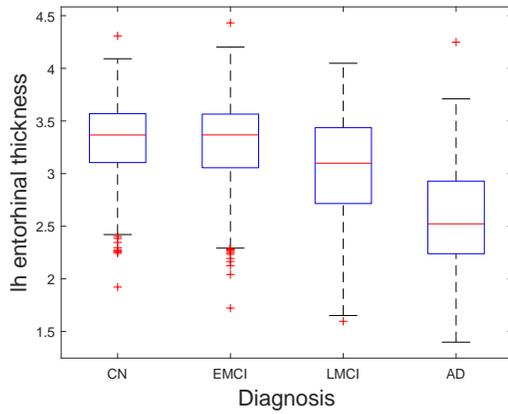
(f)



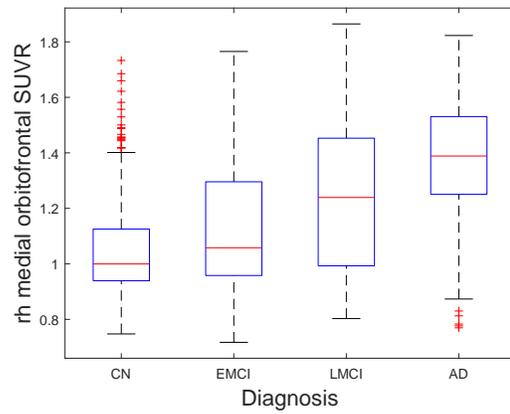
(g)



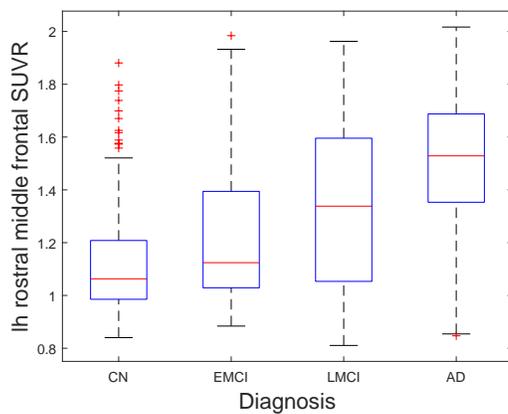
(h)



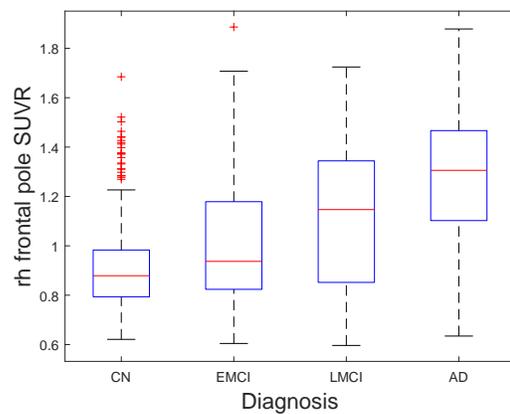
(i)



(j)



(k)



(l)

Figure 4.6: Boxplot of different features for (a)-(d): CN vs. EMCI, (e)-(h): CN vs. LMCI, and (i)-(l): CN vs. AD.

To assess the performance of the proposed method, we compared our results with the SVM classifier using the same kernel and the same features. The results as provided in Table 4.3, indicate that GP with linear kernel does not provide higher accuracy in comparison to the GP with RBF in most groups except for CN vs. LMCI. GP with the linear kernel is more successful to detect LMCI in group of CN vs. LMCI; however, the computation time for this method is significantly higher than the other methods as can be seen from Table 4.4.

Table 4.3: Performance Comparison of the Gaussian classifier with SVM using the same kernel and the same feature including the MRI, PET, and DI. (RBF: radial basis function kernel, and L-K: linear kernel)

	CN vs. EMCI			CN vs. LMCI			CN vs. AD		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
SVM (RBF-K)	75.6	78.9	70.6	76.9	69.3	85.7	91.2	90.1	93.9
GP (RBF-K)	78.8	81.3	76.8	79.8	70.2	89.9	94.7	92.3	95.9
SVM (L-K)	69.4	71.2	67.4	78.7	70.6	82.8	92.6	89.4	93.7
GP (L-K)	68.7	67.7	75.0	81.5	76.3	86.6	91.5	91.3	92.7
	EMCI vs. LMCI			EMCI vs. AD			LMCI vs. AD		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
SVM (RBF-K)	70.1	80.0	60.0	85.5	90.5	80.8	75.3	73.3	75.6
GP (RBF-K)	73.2	81.2	69.9	88.1	92.8	87.4	77.1	79.9	75.9
SVM (L-K)	72.1	79.1	68.3	79.2	84.8	67.8	80.2	80.1	80.7
GP (L-K)	72.0	78.3	67.9	81.4	81.0	83.1	70.7	70.9	73.8

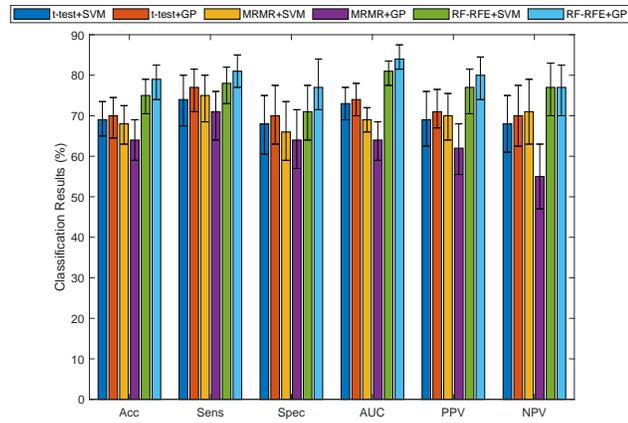
Table 4.4: Computation time in second for SVM and GP using linear and RBF kernels (Variables are represented in seconds)

	CN vs. EMCI	CN vs. LMCI	CN vs. AD
	SVM (RBF-K)	96.58	74.52
GP (RBF-K)	44.48	28.8	18.3
SVM (L-K)	15.21	11.21	22.08
GP (L-K)	2453.97	4963.02	915.47
	EMCI vs. LMCI	EMCI vs. AD	LMCI vs. AD
	SVM (RBF-K)	90.12	67.69
GP (RBF-K)	38.22	27.6	25.18
SVM (L-K)	15.77	19.62	10.62
GP (L-K)	1756.32	1833.21	729.53

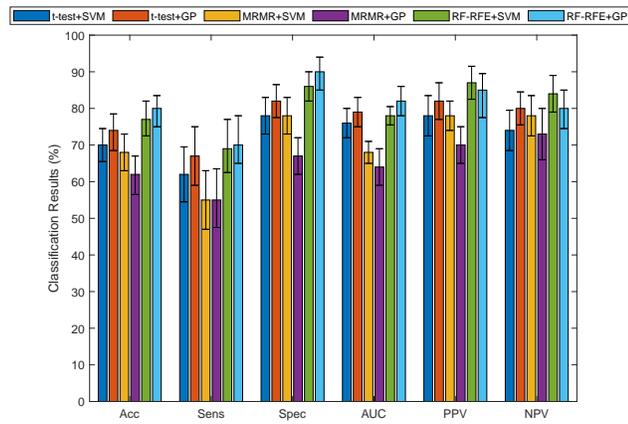
In order to assess the performance of the proposed method, the classification results of some well-established feature selection methods using SVM and GP classifiers for 3 binary classifications of CN vs. EMCI, CN vs. LMCI, and CN vs. AD are plotted in Figure 4.7. Combination of t-test with GP classifier and random forest recursive feature elimination feature selection with GP and SVM classifiers have higher accuracy in comparison to the other combination algorithms.

4.5 Discussion

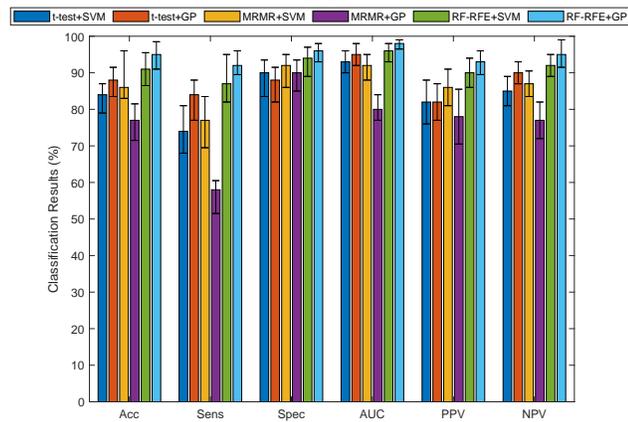
In this study, we evaluated a machine learning algorithm based on the Gaussian process for the delineation of the challenging EMCI from the CN group. The similarity of the SVM and Gaussian process is in using the covariance kernel; however, the maximum margin approach is distinct in the SVM approach. The classification results of GP and SVM are not statistically significant; however, GP provides the predicted probability of the labels which could be beneficial in the clinical investigation while SVM provides the binary predicted labels. For example, the small predicted probability for a subject will suggest more tests are required in a clinical setting. Moreover, temporal atrophy seems to be more relevant for the CN vs. LMCI binary classification rather than for the more challenging CN vs. EMCI. This indicates that memory deterioration of the medial temporal lobe mostly occurs at the transition from EMCI to the LMCI stage. In addition, we observed that the most important variables are often selected from the left hemisphere of the brain may suggest that more deterioration has taken place on the left hemisphere than the right hemisphere more, specifically in the transition from EMCI to LMCI although the right/left-handed information of the participants is not available to make a strong statement.



(a)



(b)



(c)

Figure 4.7: Classification results with 95% confidence interval for combination of different feature selection using SVM and GP classifiers for the most important classifications of (a) CN vs. EMCI, (b) CN vs. LMCI, and (c) CN vs. AD. (MRMR: minimum redundancy maximum relevance, RF-RFE: random forest recursive feature elimination)

So far, only a limited number of studies have considered EMCI and LMCI groups [8, 53–56, 58, 59, 61]. Table 4.5 lists all the classification results from previous studies which are compared to the proposed method. One advantage of the proposed method over the previously reported methods is that the results offer both higher accuracy and higher sensitivity values for most classification groups although the confidence intervals have not been considered in those studies for full comparison purposes.

Table 4.5: Accuracy (Acc), sensitivity (Sen), and specificity (Spe) of the Gaussian classifier comparing to the previous works. (dMRI: diffusion magnetic resonance imaging, fMRI: functional magnetic resonance imaging, and DI: demographic information)

		CN vs. EMCI			CN vs. LMCI			CN vs. AD		
	Modality	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
[59]	MRI	56	52	60	59	52	64	84	73	89
[53]	dMRI	59.2	-	-	62.8	-	-	78.2	-	-
[56]	MRI	65	61	69	-	-	-	86	86	85
[61]	fMRI	66.0	71.4	64.1	-	-	-	93.8	92.8	95.7
[58]	fMRI	72.8	78.3	67.1	78.6	82.5	72.2	88.9	91.7	85.7
proposed	MRI+PET+DI	78.8	81.3	76.8	79.8	70.2	89.9	94.7	92.3	95.9
		EMCI vs. LMCI			EMCI vs. AD			LMCI vs. AD		
	Modality	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
[59]	MRI	63	62	66	81	70	82	67	58	73
[53]	dMRI	63.4	-	-	-	-	-	-	-	-
[56]	MRI	-	-	-	-	-	-	-	-	-
[61]	fMRI	-	-	-	-	-	-	-	-	-
[58]	fMRI	-	-	-	-	-	-	-	-	-
proposed	MRI+PET+DI	73.2	81.2	69.9	88.1	92.8	87.4	77.1	79.9	75.9

CHAPTER 5

PREDICTION OF CONVERSION FROM NORMAL COGNITION AND MILD COGNITIVE IMPAIRMENT TO ALZHEIMER'S DISEASE

5.1 Goal

This study aims to examine disease progression in the early stage using longitudinal data. Longitudinal analysis of multimodal neuroimaging data is essential for understanding Alzheimer's disease and its progression as a function of the different risk factors. Prediction of progression from mild cognitive impairment (MCI) to AD is widely investigated; however, the conversion from cognitively normal (CN) to MCI and on to AD is largely unexplored. Identification of individuals with normal cognition that are likely to progress to MCI or AD over time will improve the planning and efficacy of any treatment in clinical trials. Therefore, we investigated conversion from both groups of CN and MCI by predicting cognitive tests including Alzheimer's disease assessment scale cognitive subscale (ADAS-Cog), Mini-mental state examination (MMSE), and Rey's auditory verbal learning (RAVLT) that have been designed and used as important criteria to evaluate the cognitive status of patients. In this study, we formulated the prediction of disease progression as a multitask regression problem by considering a task as the prediction of the cognitive score at each time point and multiple prediction tasks across all available time points simultaneously to capture the temporal smoothness of the model through an undirected dependency graph for all tasks. The proposed model learns subject's trajectories of Magnetic Resonance Imaging (MRI) features, Cerebrospinal fluid (CSF), Fluorodeoxyglucose (FDG)-PET, and Apolipoprotein E (APOE) gene with a multitask approach for every single modality and fuse the results to predict the aforementioned neuropsychological tests in a longitudinal study with a 4-year duration. In addition, we investigated the association between brain structural patterns changes in disease progression and observed that white

matter volume of the left hippocampus, cortical thickness average of left middle temporal, and right entorhinal play significant roles in predicting the cognitive scores. We also observed alternations within the insula in the conversion from MCI and normal cognition to AD.

5.2 Material and Methods

5.2.1 Participants and Data Acquisition

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database. The multimodal data in this study used longitudinal information from 779 subjects in a 4-year time window which were categorized into 3 groups: cognitively normal (CN), mild cognitive impairment (MCI), and Alzheimer’s disease (AD) at baseline (BL or T1). These subjects are subsequently divided into 5 categories of CN stable (CNs), CN converter (CNc), MCI stable (MCIs), MCI converter (MCIC), and AD in context to the next 5 time points (T6, T12, T24, T36, and T48), with 6, 12, 36, and 48 being the number of months after baseline. This is based on the conversion or stability of the subject’s cognition status within the 4-year duration. In this study, subjects that were diagnosed as CN/MCI at baseline and the status has not been changed during the 4 years are considered as CN/MCI stable. It should be mentioned that the subjects are considered converter if they progressed to either MCI or AD within the 4 years, and are considered as stable even if they eventually converted after the 4 years. All subjects that have conversions from MCI to CN or from AD to MCI are excluded in this study. Demographic information of the subjects is provided in Table 5.1. Among cognitive scores, 4 neuropsychological tests of ADAS, MMSE, and RAVLT including RAVLT immediate (RAVLT-Im) and RAVLT percent forgetting (RAVL-PF) were

selected for predictive modeling. The demographic information of participants based on converted groups is also provided in Table 5.2. In addition, evolution patterns of the different groups categorized in Table 5.2 over the 6 available time points are depicted in Figure 5.1 which shows average changes of cognitive scores in progression trajectory.

Table 5.1: Participants Demographic information at baseline. Values are represented as mean (standard deviation)

	CN	MCI	AD
Subjects	223	394	162
Age	74.51(5.29)	73.24(7.29)	74.03(7.91)
CDR	0.03(0.13)	1.51(0.83)	4.32(1.66)
MMSE	29.13(1.02)	27.37(1.75)	23.22(2.01)
ADAS	8.94(4.19)	17.42(6.52)	28.41(7.94)
RAVLT-Im	44.99(9.26)	32.66(10.06)	23.22(7.78)
RAVLT-PF	33.32(27.81)	64.08(32.19)	87.67(22.73)

Table 5.2: Participants Demographic information at baseline. Values are represented as mean (standard deviation)

	CNs	CNc	MCIs	MCIc	AD
Subjects	203	20	265	129	162
Age	74.32(5.12)	76.95(5.16)	73.16(7.37)	72.94(7.03)	74.03(7.91)
CDR	0.08(0.30)	0.52(1.02)	1.46(0.97)	3.36(2.23)	4.32(1.66)
MMSE	29.13(1.12)	28.47(1.56)	27.58(2.11)	24.77(3.75)	23.22(2.01)
ADAS	8.33(4.10)	12.71(6.10)	15.67(6.88)	25.28(8.76)	28.41(7.94)
RAVLT-Im	44.91(9.92)	38.29(10.10)	34.38(10.89)	24.08(7.90)	23.22(7.78)
RAVLT-PF	32.66(27.14)	47.64(31.49)	59.60(40.12)	83.74(39.21)	87.67(22.73)

5.2.2 Cognitive tests

MMSE is one of the most well-known tests that was designed to monitor the development of overall cognitive impairment [125]. The scale of this score is 0 to 30 which 20 to 24 is considered mild dementia, 13 to 20 indicates moderate dementia, and less than 12

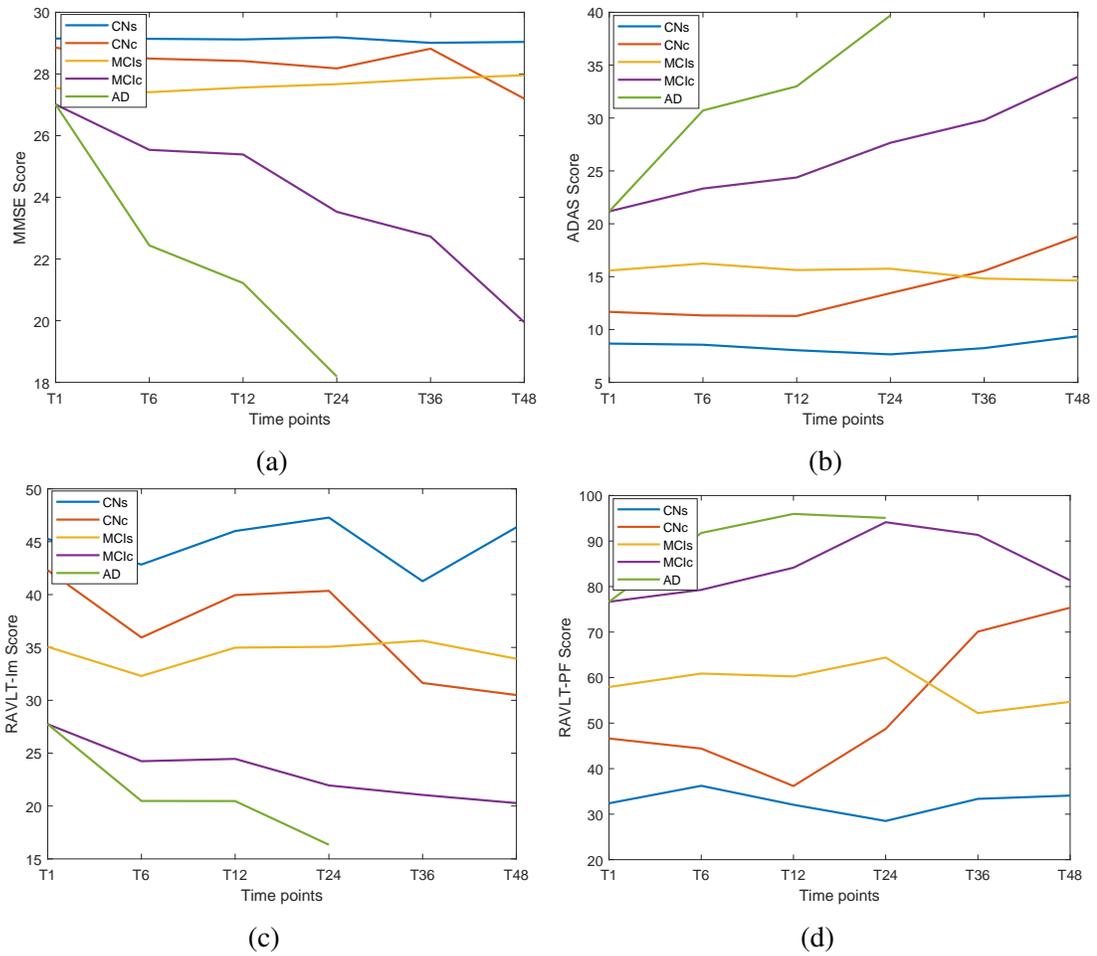


Figure 5.1: Average changes of cognitive scores for the 5 different groups of CNs, CNc, MCIs, MCIc, and AD across the 6 time points for method for (a) MMSE (b), ADAS (c), RAVLT-Im, and (d) RAVLT-PF.

suggest severe dementia; however, in many studies, MMSE scores of 0-10, 11-20, 21-25, 26-29, and 30 have been considered as a projection of CDR of 3 or above, 2, 1, 0.5, and 0 indicating severe, moderate, questionable, and normal cognition, respectively [126].

ADAS-cog which can be considered as a gold standard in antideementia treatments was developed originally for later stages of dementia where cognitive decline is more severe; however, it is often used in the earlier stage of MCI [127]. The ADAS-11 score includes 11

tasks that evaluate domains of language, ideational and constructional praxis (assembling parts into a structure), word recognition, remembering test instructions, speaking, and language word-finding difficulty. On the other hand, ADAS-13 was designed later with better performance for disease progression in the early stage, which combines the 11 tasks of ADAS-11 with a delayed word recall as well as a maze task added. Thus, for practical purposes, only ADAS-13 was considered here as ADAS test score.

RAVLT is another powerful test to assess verbal learning and episodic memory during the progression of Alzheimer's disease. RAVLT procedure is based on repeated tasks of presenting 15 words in consecutive trials for the subjects and ask them to recall as many words as they can remember for those tasks at each trial [128].

5.2.3 Neuroimaging

In this study multimodal neuroimaging of MRI, FDG-PET, and CSF along with APOE $\epsilon 4$ allele were used. The number of MRI measurements are 319 including 275 cortical features (left/right white matter volume, cortical volume, surface area, cortical thickness, and standard deviation of thickness plus total intracranial volume (ICV)) and 44 subcortical features including left/right subcortical volumes of subcortical brain regions. The FDG-PET biomarkers determine the metabolic changes in the brain. CSF biomarker includes 3 features of phosphorylated tau protein (p-Tau), amyloid- β (A- β), and total tau protein (t-Tau). Tau plays an important role in adjusting axonal transport, microtubule dynamics, and neurite outgrowth which lead to clinical diagnostic settings of Alzheimer's disease [129–131]. Finally, APOE $\epsilon 4$ allele is responsible for carrying lipids through the bloodstream which is considered as a major genetic risk factor for AD [132, 133]. On the

other hand, as mentioned before all subjects have measurements at the baseline or T1 but may not have all neuroimaging measures for the next five time points of T6, T12, T24, T36, and T48. Table 5.3 Shows the number of observations for the different modalities at different time points. We can see from Table 5.3 the extent of missing data especially and expectedly for the PET and CSF, which highlights the importance and necessity of managing the missing values.

Table 5.3: Number of observations for different modalities at different time points of T1-T48

Modality	T1	T6	T12	T24	T36	T48
MRI	779	695	654	503	238	144
PET	521	209	194	286	81	46
CSF	501	4	181	143	35	50
APOE	779	695	654	503	238	144

5.2.4 Fused Gaussian Sparse Group Lasso (FGSGL)

In longitudinal study of Alzheimer’s disease, predicting the cognitive scores at any time point is a regression problem that can be developed as a multitask regression problem by predicting multiple tests or predicting a test at multiple time points. In this study a multitask learning approach has been applied for the prediction of any of the considered cognitive scores for all time points T1 through T48. Let us consider the input matrix as $X_t = \{x_1, x_2, \dots, x_N\}$ and target matrix or neuropsychological test score as $y_t = \{y_1, y_2, \dots, y_N\}$, therefore, $X_t \in R^{N \times F}$ and $y_t \in R^N$ where N is the number of observations and F is the number of features of neuroimaging at the time of $t = 1, 2, \dots, T$. It should be noted that all vectors are defined with lowercase letters, and matrices are defined with uppercase letters throughout this article. If the regression parameters across all tasks are considered as $\Phi \in R^{F \times T}$ matrix, then $\varphi \in R^F$ denotes the column of regression parameters of the task

at the time t . $W_t = \{w_1, w_2, \dots, w_T\}$ is the weight matrix at all time points. A kernel-based smoothing approach [134] which is used to local smoothing in order to minimize the regression error at each time point is associated with the task t and neighbor φ_t . Thus, the approximation model can be determined as below:

$$\hat{\varphi}_t = \sum_{\substack{r=1 \\ r \neq t}}^T w_{r,t} \varphi_r, \quad t = 1, 2, \dots, T \quad (5.1)$$

$$\text{where } w_{r,t} = \frac{K(\frac{r-t}{\sigma})}{\sum_{\substack{r=1 \\ r \neq t}}^T K(\frac{r-t}{\sigma})}, \quad r = 1, 2, \dots, T, r \neq t$$

Here σ is the bandwidth and K is the kernel matrix using the Gaussian kernel as

$$K = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2}\right) \quad (5.2)$$

In Equation (5.1), the weights are defined by the Gaussian kernel where its bandwidth needs to be determined. A small value of σ leads to quick decay of the Gaussian curve, and vice versa the larger value is to allow for the curve to decay gradually. We determined $\sigma = 14$ as an appropriate empirical value to be used in this study.

On the other hand, the fused aspect of the model is obtained by adding sparsity on the matrix of residuals [135]. The fused penalty or the transformation matrix as used in this study can be defined as $G \in R^{(T \times T)}$ in the term of $P = \Phi G$ as follow

$$\begin{bmatrix} \rho_1 & \rho_2 & \cdots & \rho_T \end{bmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_T \end{bmatrix} \begin{bmatrix} I & -w_{|t-r|}I & -w_{|t-r|}I & \vdots & -w_{|t-r|}I \\ -w_{|t-r|}I & I & \vdots & \vdots & -w_{|t-r|}I \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -w_{|t-r|}I & -w_{|t-r|}I & -w_{|t-r|}I & \cdots & I \end{bmatrix} \quad (5.3)$$

The matrix of G includes the weights $w_{t,r} = w_{|t-r|}$ demonstrating the edges between the nodes t and r . Therefore, the solution for the multitask problem is to solve the following unconstrained optimization equation:

$$\min_{\Phi, P} \sum_{t=1}^T \|y_t - X_t \varphi_t\|^2 + H_{\beta_2}^{\beta_1}(\Phi) + \beta_3 \|P\|_1, \quad \rho_t = \varphi_t - \sum_{\substack{r=1 \\ r \neq t}}^T w_{r,t} \varphi_r \quad (5.4)$$

where the columns of residuals ρ_t creates the matrix of residuals $P \in R^{F \times T}$, the β_1 , β_2 , and β_3 are the regularization parameters, and $H_{\beta_2}^{\beta_1}(\Phi) = \beta_1 \|(\Phi)\|_1 + \beta_2 \|(\Phi)\|_{2,1}$ denotes the combination of penalties of the lasso and the group lasso. The group lasso defined as $\|(\varphi)\|_{2,1} = \sum_{i=1}^F \|\varphi_i\|$ considers the groups across all time points for each variable i , which allows sharing a common set of variables at each time point.

In the next step, an alternating direction method of multiplier (ADMM) is used to solve the unconstrained optimization problem in Equation (5.4) which is difficult to be optimized directly. To this purpose consider the formulation as a linear constrained optimization problem as follow

$$\min_{\Phi, P, \Gamma, \Theta} \sum_{t=1}^T \frac{1}{2} \|y_t - X_t \varphi_t\|^2 + H_{\beta_2}^{\beta_1}(\Theta) + \beta_3 \|\Gamma\|_1 \quad (5.5)$$

here Θ and Γ are the feasible sets where $\Phi - \Theta = 0$, $P - \Gamma = 0$, and the column of residuals in the feasible set is defined as $\rho_t = \varphi_t - \sum_{\substack{r=1 \\ r \neq t}}^T w_{r,t} \varphi_r$.

The augmented Lagrangian can be determined as follows:

$$\begin{aligned}
L_\delta(\Phi, P, \Theta, \Gamma, A, M, N) = & \sum_{t=1}^T \frac{1}{2} \|y_t - X_t \varphi_t\|^2 + H_{\beta_2}^{\beta_1}(\Theta) + \beta_3 \|\Gamma\|_1 + \text{tr}(A^T(\Phi - \Theta)) \\
& + \frac{\delta}{2} \|\Phi - \Theta\|^2 + \sum_{t=1}^T \left\{ \mu_t^T \left(\varphi_t - \sum_{\substack{r=1 \\ r \neq t}}^T w_{|t-r|} \varphi_r - \rho_t \right) + \frac{\delta}{2} \left\| \varphi_t - \sum_{\substack{r=1 \\ r \neq t}}^T w_{|t-r|} \varphi_r - \rho_t \right\|^2 \right\} \\
& + \text{tr}(N^T(P - \Gamma)) + \frac{\delta}{2} \|P - \Gamma\|^2
\end{aligned} \tag{5.6}$$

where $A, M, N \in R^{F \times T}$ are Lagrangian multipliers associated with $\Phi - \Theta = 0$, $\varphi_t - \sum_{\substack{r=1 \\ r \neq t}}^T w_{|t-r|} \varphi_r - \rho_t = 0$, and $P - \Gamma = 0$ constraints, including the vectors of α , μ , and ν , respectively. In addition, $\text{tr}(\cdot)$ shows the trace of a matrix and δ is the penalty parameter which is used to define the dual ascent step size in ADMM [136].

If we define $\Phi(q)$ as in Equation (5.7) and considering q_t^k as the gradient with respect to φ_t , then the regression parameter matrix can be calculated through using Algorithm 1 as follow:

$$q(\Phi) = \frac{\delta}{2} \sum_{t=1}^T \left\| \varphi_t - \sum_{\substack{r=1 \\ r \neq t}}^T w_{|t-r|} \varphi_r - \rho_t \right\|^2 \tag{5.7}$$

Finally, the regression matrix Φ is obtained with N number of observation at $t = T$ time point samples. Therefore, M modality-specific regression matrices generate the M primary prediction of each cognitive score of \hat{y} using $\hat{y}_M^t = X_M^t \times \Phi_M^t$.

Algorithm 1: Steps to update primal and dual variables for ADMM

- Determine the initial variables of A^k , M^k , N^k , Θ^k , Γ^k , P^k , and Φ^k
 - Repeat the following steps until termination criterion is satisfied
 - $\varphi_t^{k+1} \leftarrow \underset{\varphi_t}{\operatorname{argmin}} \frac{\delta}{2} \|y_t - X_t \varphi_t\|^2 + (\alpha_t^k)^T \varphi_t + \frac{\delta}{2} \|\varphi_t - \theta_t^k\|^2 + (\mu_t^k + q_t^k)^T \varphi_t + \frac{\delta}{2} \|\varphi_t - \varphi_t^k\|^2$
 - $\rho_t^{k+1} \leftarrow \underset{\rho_t}{\operatorname{argmin}} \frac{\delta}{2} \|\rho_t - \{\gamma_t^k + \varphi_t^{k+1} - \sum_{r \neq t}^T w_{|t-r|} \varphi_r^{k+1}\}\|^2 - (\mu_t^k + \nu_t^k)^T \rho_t$
 - $\Theta^{k+1} \leftarrow \operatorname{sign}(\Phi^{k+1} + \frac{1}{\delta} A^k) \max(|\Phi^{k+1} + \frac{1}{\delta} A^k| - \frac{\beta_1}{\delta}, 0)$
 - $\Theta^{k+1} \leftarrow \frac{\max\{\|\Theta^{k+1}\|_2 - \frac{\beta_2}{\delta}, 0\}}{\|\Theta^{k+1}\|_2}$
 - $\Gamma^{k+1} \leftarrow \operatorname{sign}(P^{k+1} + \frac{1}{\delta} N^k) \max(|P^{k+1} + \frac{1}{\delta} N^k| - \frac{\beta_3}{\delta}, 0)$
 - $A^{k+1} \leftarrow A^k + \delta(\Phi^{k+1} - \Theta^{k+1})$
 - $\mu_t^{k+1} \leftarrow \mu_t^k + \delta(\varphi_t^{k+1} - \sum_{r \neq t}^T w_{|t-r|} \varphi_r^{k+1} - \rho_t^{k+1})$
 - $N^{k+1} \leftarrow N^k + \delta(P^{k+1} - \Gamma^{k+1})$
 - $A^k = A^{k+1}, N^k = N^{k+1}$
 - $\varphi_t^k = \varphi_t^{k+1}, \mu_t^k = \mu_t^{k+1}, \theta_t^k = \theta_t^{k+1}$
 - $\Phi \leftarrow \Theta$
-

5.2.5 Fusion Algorithm

In the second step, we need to ensemble the primary predictions of specific-modalities which was calculated using the FGSGGL, as described in section 5.2.4, separately. An ensemble technique is used to combine the predictions from multiple classification or regression algorithms to make more accurate predictions that could be achieved from any learning algorithm alone. Here a least-squares boosting (LS-Boost) was used to combine multiple weak learners into one strong learner in order to fuse the specific-modalities predictions. The algorithm uses multiple decision tree regressors to train the network sequentially with respect to residual errors made by the previous regressor. Therefore, missing data can be handled using the decision trees for training the network as well as minimizing the error using sequential training. If there would be any missing data, the algorithm selects a new split with input data which is called surrogate split, that is not to involve the missing data in the training process.

Here the input matrix is defined as $\hat{y}^t = [\hat{y}_1^t, \hat{y}_2^t, \dots, \hat{y}_M^t]$ from the last section. For simplicity in notation, we consider the input matrix of $X = [x_1, x_2, \dots, x_M] \in R^{N \times 1}$ where N is the number of observations and M is the number of modalities that are going to be fused at each time point. Therefore, the response vector of the cognitive score can be defined as $y \in R^{N \times 1}$. We assume that X is normalized to have zero mean and unit l_2 -norm and y has also zero mean. Therefore, the estimated response vector is calculated by $X\lambda$ with the residuals of $k = y - X\lambda$. The following Algorithm provides the steps to find the regression coefficient vector of λ [137].

Algorithm 2: The two-step algorithm for least square boosting (LS-Boost)

- Fix number of iterations $t > 1$ and the learning rate of $\alpha > 0$
- Determine the initial values as $\hat{k}^0 = y$, $\hat{\lambda}^0 = 0$, and $m = 0$
- For $0 \leq m \leq t$ repeat the following
 - Find the covariates of j_m and \tilde{u}
 - * $\tilde{u} \leftarrow \underset{u \in R}{\operatorname{argmin}} (\sum_{i=1}^N (\hat{k}_i^m - x_{ir}u)^2)$ for $r = 1, 2, \dots, M$
 - * $j_m \leftarrow \underset{0 \leq r \leq M}{\operatorname{argmin}} \sum_{i=1}^N (\hat{k}_i^m - x_{ir}\tilde{u})^2$
 - Update the regression coefficients and residuals as
 - * $\hat{k}^{m+1} \leftarrow \hat{k}^m - \alpha X_{j_m} \tilde{u}_{j_m}$
 - * $\hat{\lambda}_{j_m}^{m+1} \leftarrow \hat{\lambda}_{j_m}^m + \alpha X_{j_m} \tilde{u}_{j_m}$
 - * $\hat{\lambda}_j^{m+1} \leftarrow \hat{\lambda}_j^m, \quad j \neq j_m$

The overall view of the two-step proposed framework is shown in Figure 5.2. The empty boxes in Multitask learning step represent missing values for primary predictions due to missing input data of PET, CSF, and APOE in comparison to MRI. Since all subjects have MRI but not necessarily PET, CSF, and APOE information, predicted cognitive scores resulted from those modalities would have missing values.

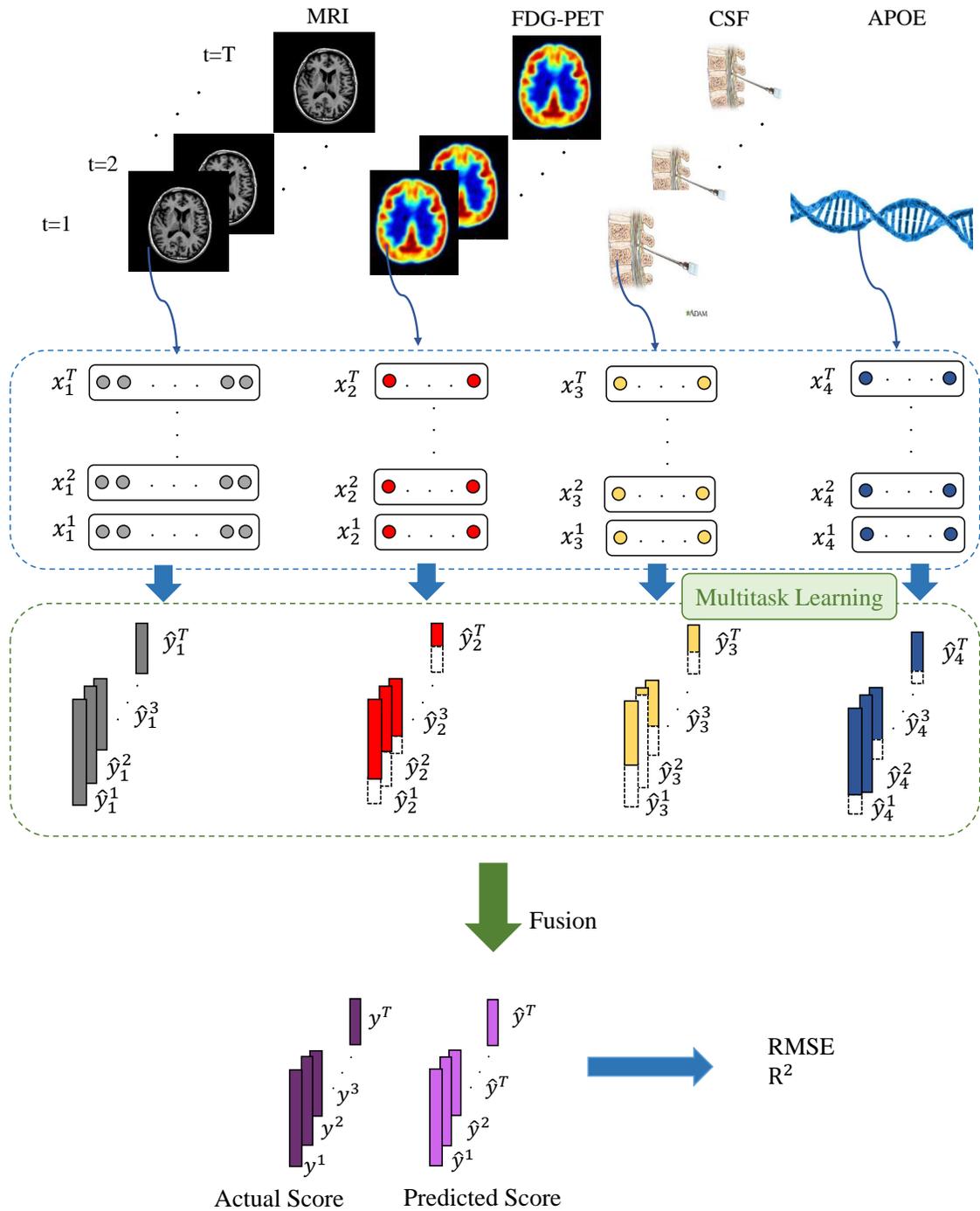


Figure 5.2: Overall view of the proposed framework to predict neuropsychological test scores. The portion highlighted in white in the predicted \hat{y}_i^t vectors signify the proportion of missing data.

5.3 Results

In this study, a 10-fold cross-validation has been used to prevent any bias in the training and testing datasets. The hyperparameters of β_1 , β_2 , and β_3 for regularization were selected from 1 to 10. In addition, the optimal number of decision trees in boosting was selected from 1 to 100. The data has been normalized to have zero mean and unit variance after splitting the data into training and testing sets. The performance of predicting the 4 cognitive scores of MMSE, ADAS, RAVLT-Im, and RAVLT-PF using multimodal data including MRI, PET, CSF, and APOE is evaluated in terms of Root Mean Square Error (RMSE) and correlation coefficient (R^2) between predicted values and actual values in the testing phase. The mean and standard deviation of RMSE obtained from each trial are averaged for 10-fold cross-validation. The final results in terms of RMSE are presented in Table 5.4. The scatter plots of predicted cognitive scores of MMSE and ADAS versus actual scores along with the correlation of R^2 are shown in Figures 5.3 and 5.4. In addition, the correlation of 0.68, 0.64, 0.66, 0.67, 0.71, and 0.73 are obtained for the prediction of RAVLT-Im scores across the 6 time points of T1 through T48, respectively. The correlation of 0.62, 0.67, 0.56, 0.69, 0.61, and 0.70 are also obtained for prediction of RAVLT-PF.

Table 5.4: Prediction performance in terms of RMSE for the 4-year time window based on multimodal neuroimaging. Values are represented as mean (standard deviation)

Tests	T1	T6	T12	T24	T36	T48
MMSE	1.69(0.15)	2.01(0.30)	1.72(0.20)	1.89(0.17)	1.94(0.29)	1.80(0.61)
ADAS	4.92(0.64)	5.56(0.53)	6.11(0.73)	5.34(0.59)	5.90(0.91)	6.69(1.13)
RAVLT-Im	7.75(0.69)	8.27(0.63)	8.16(0.87)	7.97(1.24)	8.24(0.94)	7.91(1.60)
RAVLT-PF	20.37(4.3)	24.37(5.6)	22.26(1.7)	21.29(2.5)	25.80(5.5)	23.85(4.8)

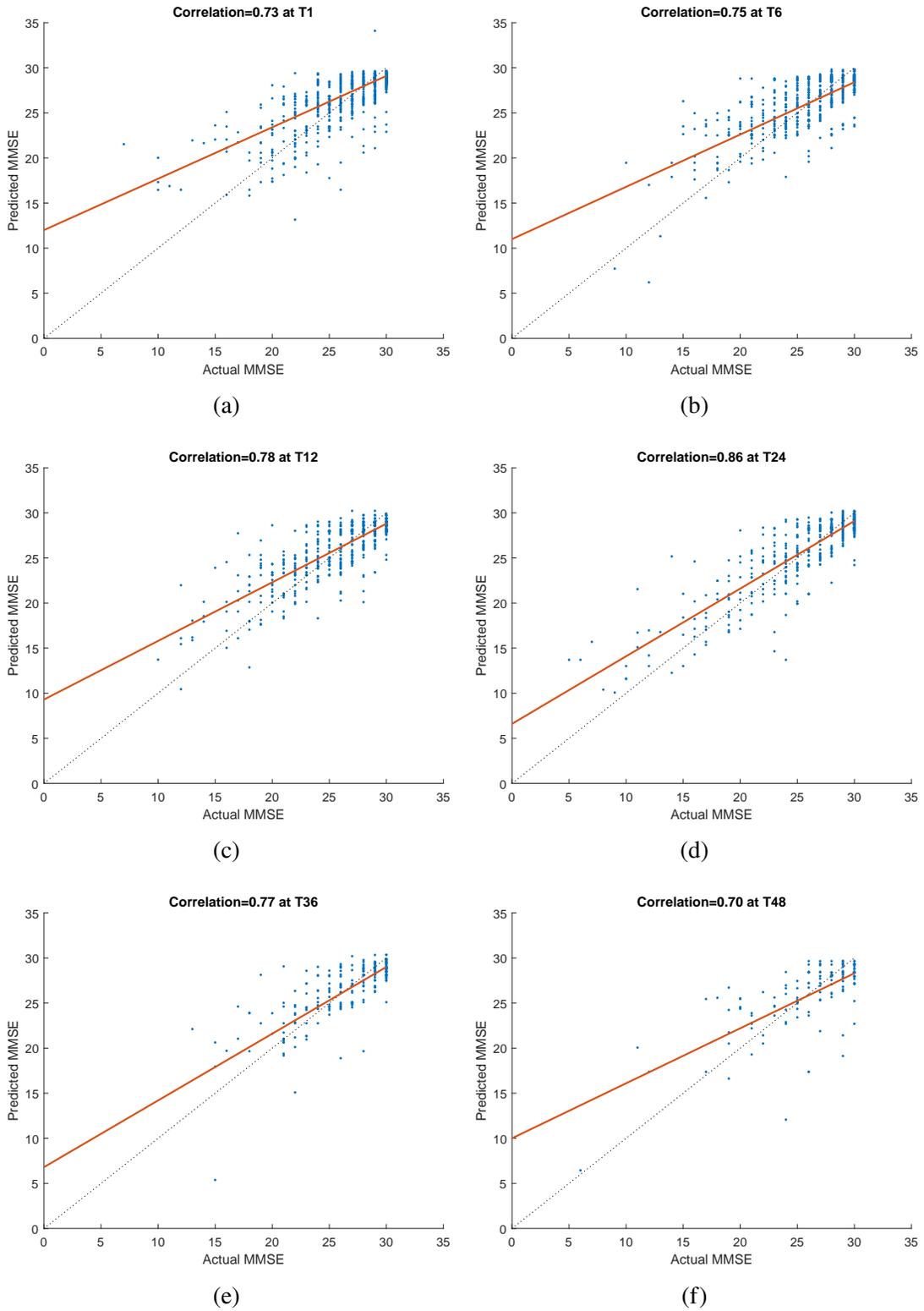


Figure 5.3: Scatter plot of predicted MMSE score vs. actual MMSE score along with the correlation coefficient at 6 time points. The orange line is regression line and the dotted gray line is the reference of perfect correlation.

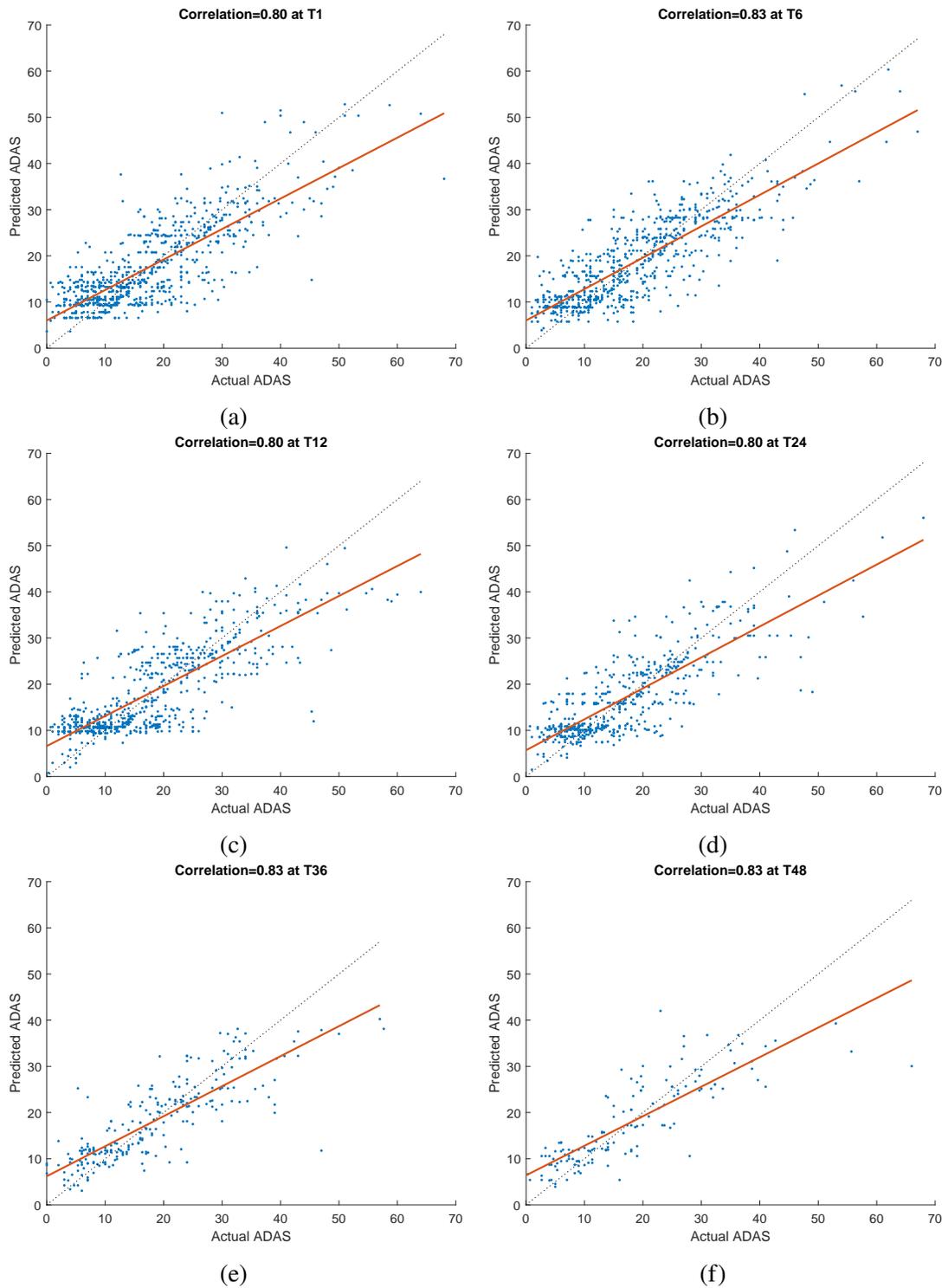


Figure 5.4: Scatter plot of predicted ADAS score vs. actual ADAS score along with the correlation coefficient at 6 time points. The orange line is regression line and the dotted gray line is the reference of perfect correlation.

As mentioned before, we considered 5 groups of CNs, CNc, MCIs, MCIc, and AD for predicting the progression from CN to MCI or AD as well as MCI to AD. Subjects were grouped as CN converter if they were diagnosed CN at baseline and then converted to MCI or AD in the next 4 years, but not converting back to CN in that time window. In the same procedure, subjects were grouped as MCI converter if they were diagnosed as MCI at baseline and converted to AD but not converting back to CN or MCI within the next 4 years. In addition, the subjects were considered CNs/MCIs if they were CN/MCI at baseline and did not convert at any of the next 5 time points even if they convert afterward. Table 5.5 show RMSE of these five categories separately for all time points.

Table 5.5: Prediction performance of MMSE and ADAS for the 4-year time window based on multimodal neuroimaging for different groups of subjects separately

MMSE	T1	T6	T12	T24	T36	T48
CNs	1.48(0.27)	1.38(0.18)	1.37(0.19)	1.38(0.18)	1.18(0.12)	1.24(0.32)
CNc	1.95(0.48)	1.87(0.10)	1.32(0.16)	1.33(0.15)	1.54(0.45)	1.26(0.88)
MCIs	1.79(0.46)	1.75(0.09)	1.98(0.16)	1.97(0.16)	1.73(0.15)	1.81(0.23)
MCIc	2.71(0.76)	2.47(0.09)	2.77(0.14)	2.93(0.39)	2.04(0.82)	2.76(0.76)
AD	3.10(0.63)	3.07(0.54)	3.32(0.73)	3.35(1.21)	-	-
ADAS	T1	T6	T12	T24	T36	T48
CNs	5.12(0.88)	5.47(0.73)	5.66(0.78)	5.00(1.47)	4.63(2.18)	4.12(2.38)
CNc	6.14(1.01)	6.16(1.77)	6.07(1.51)	4.07(1.98)	4.52(1.12)	5.35(2.26)
MCIs	4.70(0.93)	5.45(1.08)	5.65(0.83)	5.67(1.33)	5.36(2.05)	4.44(1.08)
MCIc	5.87(1.04)	5.72(1.62)	6.35(1.23)	6.80(1.12)	5.85(1.85)	6.23(2.05)
AD	6.12(1.48)	7.60(1.45)	7.95(1.56)	7.54(0.87)	-	-

Figure 5.5 provides a box plot of the distributions of predicted cognitive scores vs. actual scores for the 5 groups separately and combined. We obtained RMSE of 1.95 for normal cognition groups who later converted to MCI or AD and RMSE of 2.71 for the MCI group who converted to AD later during the 4 years.

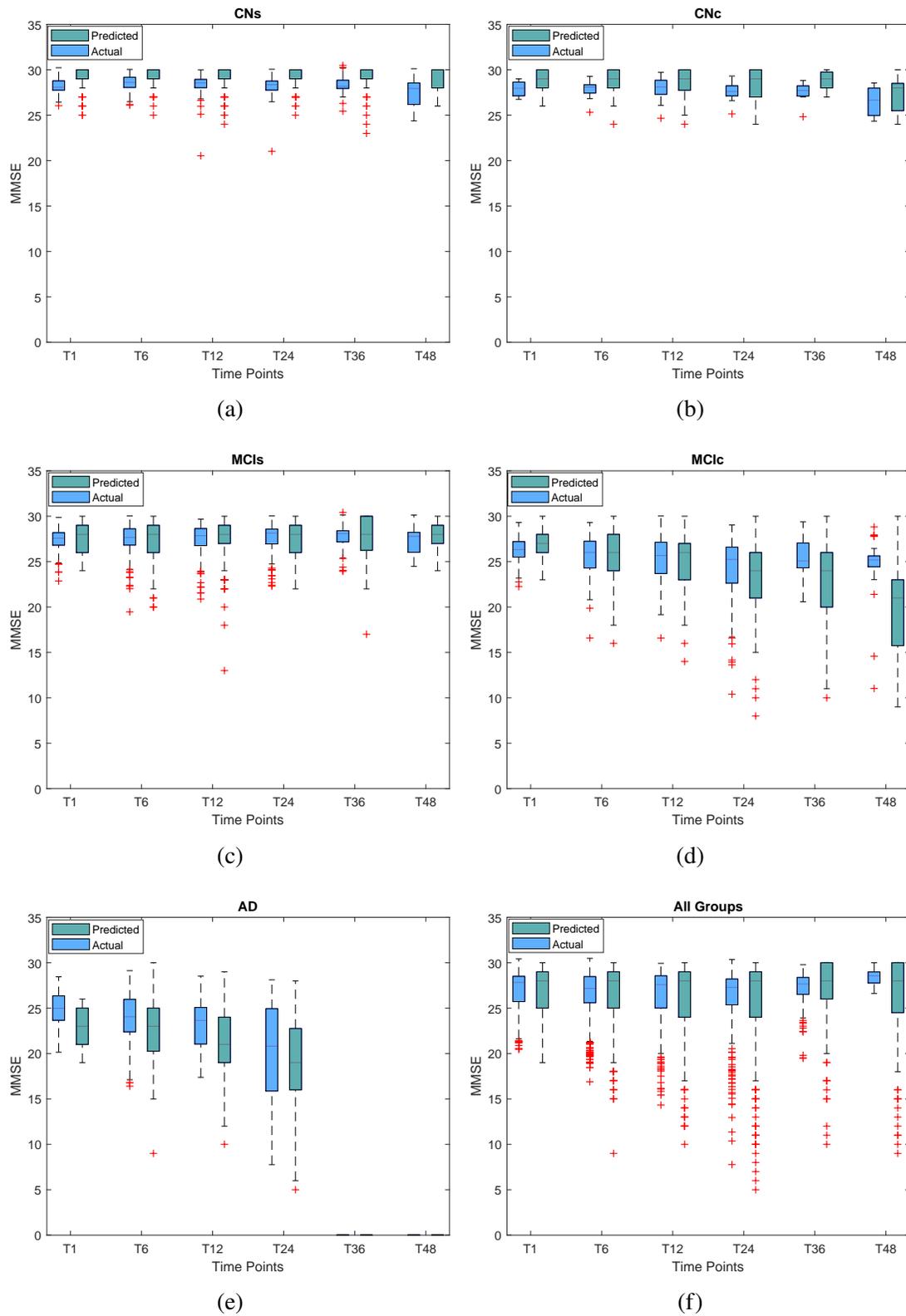


Figure 5.5: Box plot of prediction MMSE score vs. actual MMSE score using the test set based on multimodal neuroimaging for all groups of subjects separately and combined

Furthermore, Figures 5.6 and 5.7 show the delineation of CNs vs. CNc, MCIs vs. MCIc, CN vs. MCI, and CN vs. AD using the ROC curve and the area under the curve (AUC) for predicted cognitive scores vs. actual values at baseline. The predicted cognitive tests were learned from all groups in the test set and then the ROC curve along with the AUC was depicted for the selected groups. The AUC of 0.79 and 0.76 are achieved for MCIs vs. MCIc groups using the predicted MMSE and ADAS, respectively. The AUC of 0.69 and 0.72 are achieved for CNs vs. CNc groups using the predicted MMSE and ADAS, respectively. In addition, the AUC of 0.74 and 0.77 are achieved for MCIs vs. MCIc groups as well as AUC of 0.69 and 0.73 for CNs vs. CNc groups using the predicted RAVLT-Im and RAVLT-PF, respectively. The high AUC of estimated cognitive scores for these groups indicates the power of these tests to predict the conversion in the early stage.

It can be seen that the AUC for the predicted cognitive scores is slightly higher in comparison to the AUC of actual values especially for the CNs vs. CNc groups for all considered cognitive tests. These results indicate that estimated cognitive scores based on structural and functional alternations of the brain contain valuable information for early diagnosis of the disease which is different from the actual information. In order to test this hypothesis, an SVM classifier with 10-fold cross-validation was used to classify the CNs from CNc using the actual cognitive scores and combined of estimated and actual values. We achieved an average accuracy of 0.63, 0.70, 0.69, and 0.63 for MMSE, ADAS, RAVLT-Im, and RAVLT-PF, respectively using the actual values and the average accuracy of 0.69, 0.73, 0.70, and 0.71 using the combined scores which are found to be statistically significant with the null hypothesis of $P\text{-value} \leq 0.05$. These results may suggest that the actual cognitive test scores and predicted scores based on neuroimaging contain different information which may help for early detection of AD.

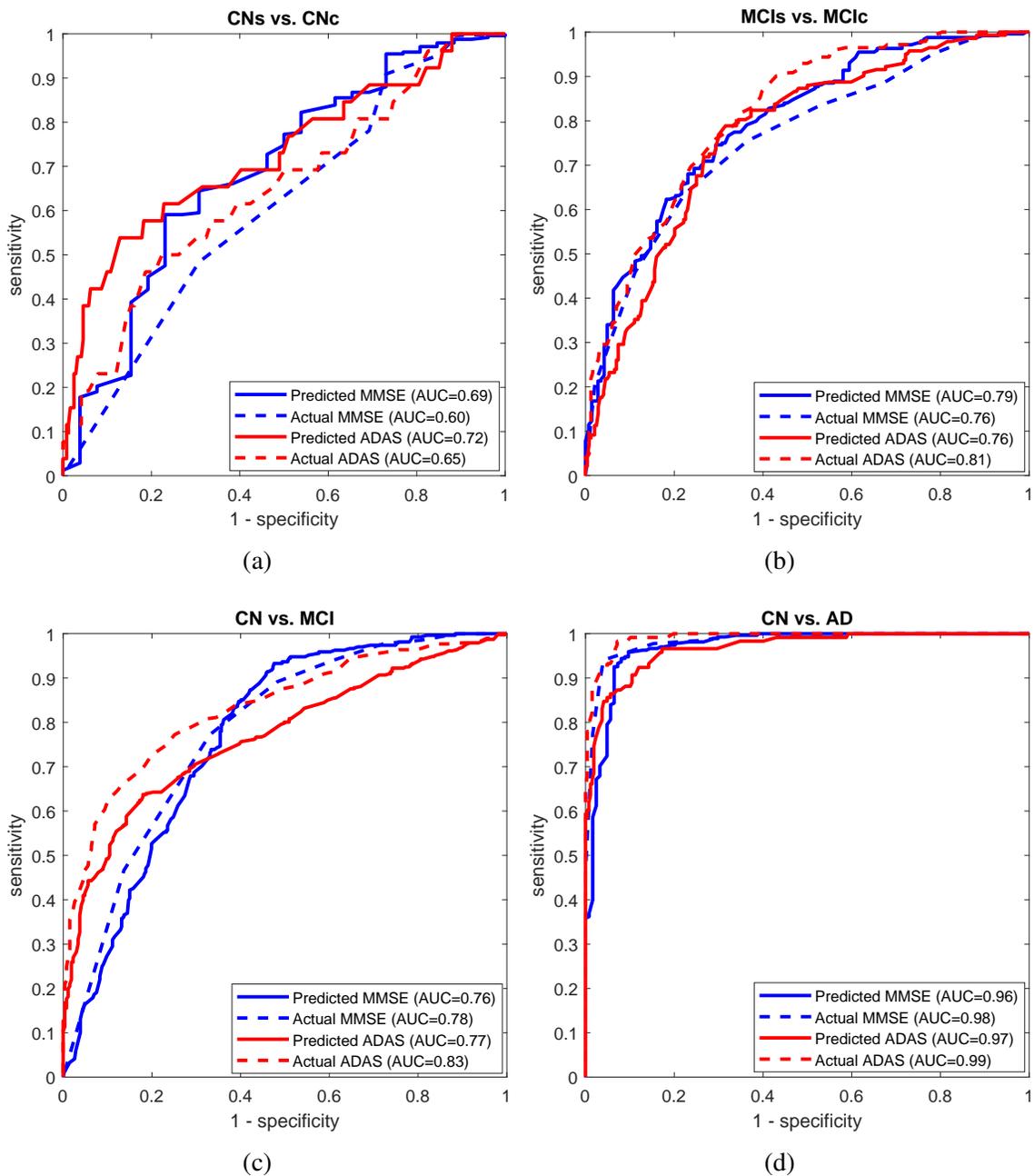


Figure 5.6: ROC curves and AUC values of group classification of CNs vs. CNc, MCIs vs. MCic, CN vs. MCI (includes both groups of stables and converters), and CN vs. AD for cognitive scores of MMSE and ADAS.

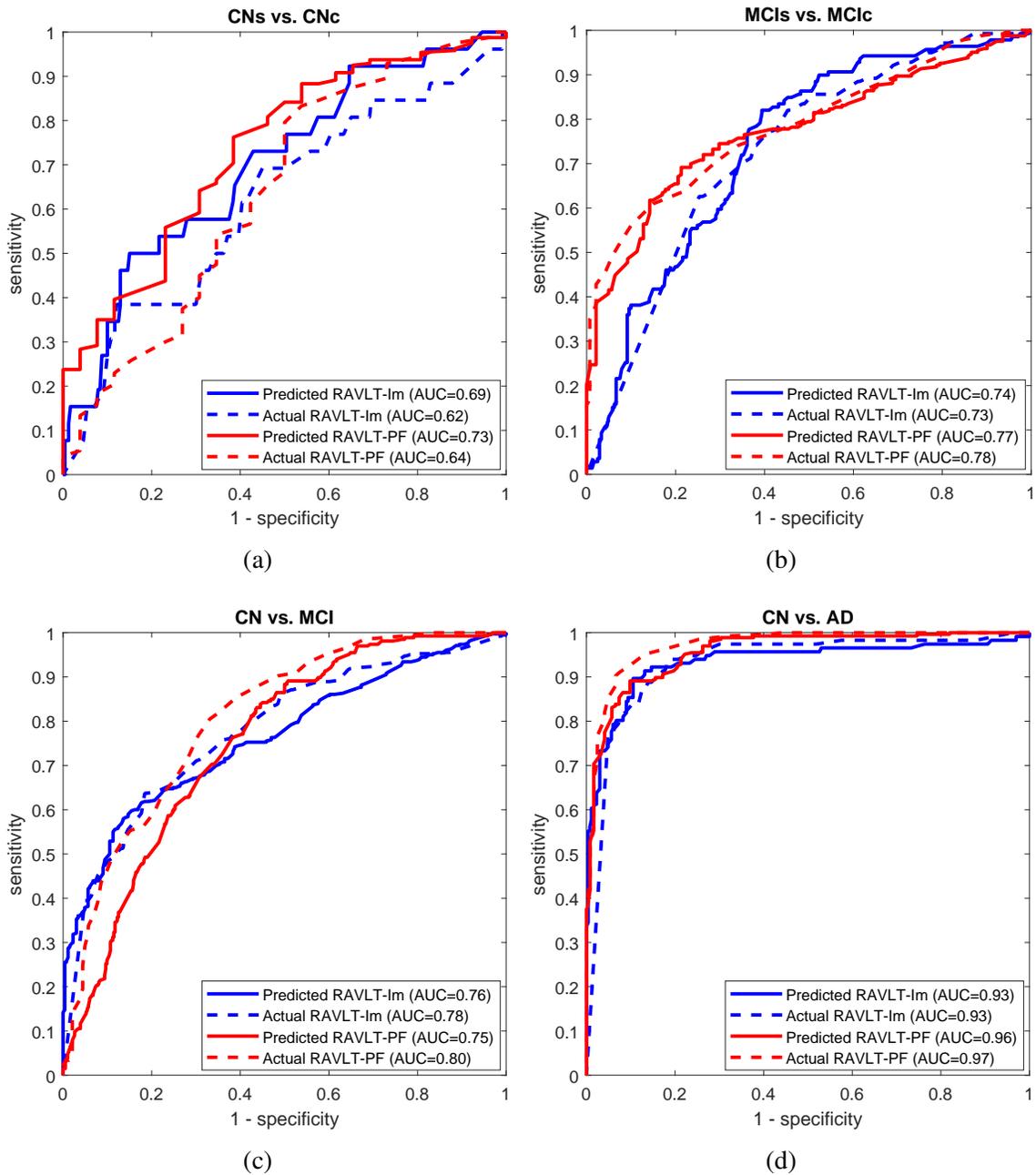


Figure 5.7: ROC curves and AUC values of group classification of CNs vs. CNc, MCIs vs. MCIC, CN vs. MCI (includes both groups of stables and converters), and CN vs. AD for cognitive scores of RAVLT-Im and RAVLT-PF.

Figure 5.8 shows the bar plot of predicted cognitive scores for a different combination of modalities. As can be seen, the RMSE is lower using only MRI modality for prediction of neuropsychological test scores in comparison to PET and CSF separately, and using multimodal neuroimaging leads to lower RMSE for all 4 cognitive tests at most of the time points. When the modalities considered (MRI, PET, CSF) are used separately, it is observed that MRI has a consistently lower RMSE in predicting cognitive scores. Evidently, the RMSE decreases when using multimodality in comparison to a single modality for predicting the cognitive scores. For almost all cases and for most time points, the more modalities or measures are combined (MRI, PET, CSF, APOE), the lower is the RMSE. However, we investigated the statistical analysis of these results using ANOVA test to check the statistical difference of RMSE and R^2 . The P-values of 0.0000, 0.2310, 0.0006, 0.0819, 0.1433, and 0.0002 for the prediction of MMSE and P-values of 0.0004, 0.0001, 0.0005, 0.0089, 0.0050, and 0.2123 for the prediction of ADAS were obtained using the single modality of MRI versus multimodal neuroimaging for the 6 time points, respectively which showed that the proposed method statistically performed best in predicting MMSE and ADAS scores only at some time points although a higher correlation and lower RMSE were achieved at all 6 time points. The same results also were observed for RAVLT-Im and RAVLT-PF prediction results. A large amount of missing data may explain this outcome since we have unbalanced observations for MRI versus PET and CSF at those time points.

The results of the proposed Gaussian-kernel-based model compared to some other regression models such as lasso, ridge, fused group lasso (FGL), temporal group lasso (TGL), and convex fused sparse group lasso regression are listed in Table 5.6. For a fair comparison, the same train and test data sets have been used for all methods.

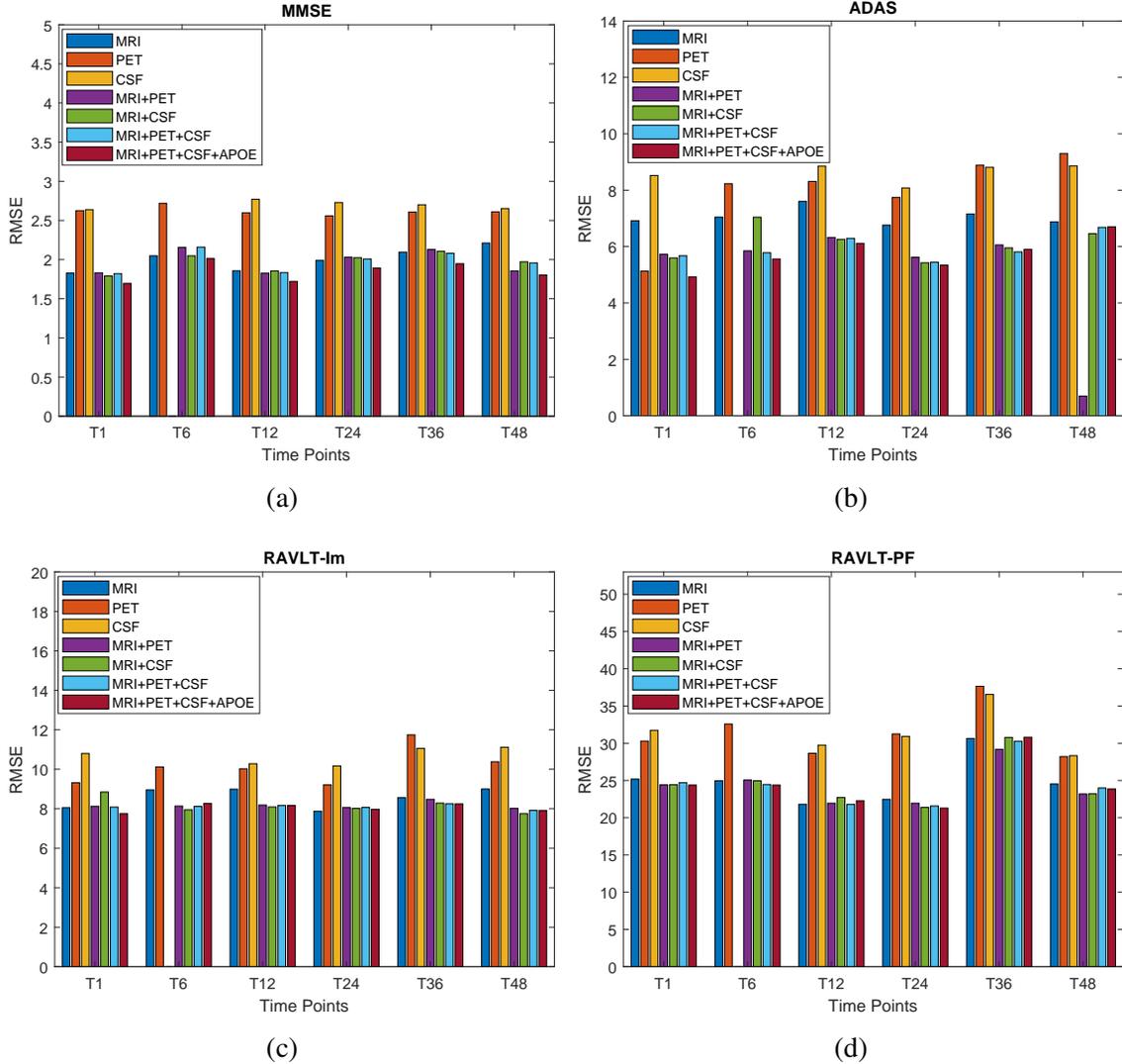


Figure 5.8: Bar plot of RMSE for different combination of modalities at different time points in 4-year time window for the cognitive tests of (a) MMSE, (b) ADAS, (c) RAVLT-Im, and (d) RAVLT-PF.

It should be mentioned that the hyperparameters of all methods were optimized in each trial with 10-time repetition using 5 fold cross-validation. Therefore, only results of MRI modality for predicting MMSE and ADAS are compared to exemplify this comparison while adhering to space constraints of this study, which we would exceed if we were to include all other modalities and for all cognitive scores. The statistical analysis

Table 5.6: Results of our proposed method compared to lasso, ridge, temporal group lasso (TGL), fused group lasso (FGL), and convex fused sparse group lasso (cFSGL) in terms of averaged RMSE (standard deviation) using MRI data to predict MMSE and ADAS. Superscript symbol of * indicates that the marked method significantly outperformed the others at that time point

MMSE	T1	T6	T12	T24	T36	T48
Ridge	3.18(0.39)	2.67(0.26)	3.05(0.45)	3.23(0.39)	3.23(0.44)	4.27(0.21)
Lasso	2.35(0.24)	2.37(0.21)	2.76(0.55)	2.95(0.35)	2.85(0.38)	4.00(0.30)
TGL	2.34(0.27)	2.04(0.26)*	2.72(0.54)	3.00(0.41)	3.02(0.36)	3.99(0.43)
FGL	3.32(3.28)	5.27(0.40)	4.21(0.41)	3.45(0.51)	2.82(0.38)	7.58(0.89)
cSFGL	2.41(0.21)	2.27(0.21)	2.75(0.44)	3.20(0.65)	2.87(0.65)	2.56(0.55)
Proposed	1.89(0.17)*	2.24(0.22)	2.03(0.25)*	1.95(0.29)*	2.09(0.32)*	2.35(0.65)
ADAS	T1	T6	T12	T24	T36	T48
Ridge	8.11(0.72)	9.86(0.98)	9.56(0.81)	10.97(1.21)	12.87(1.86)	13.27(2.65)
Lasso	7.20(0.61)	8.64(0.13)	8.23(0.81)	9.95(0.95)	8.89(1.30)	9.83(1.90)
TGL	6.98(0.64)	7.83(0.24)	7.86(0.19)	8.51(0.26)	6.85(0.88)*	9.85(0.98)
FGL	10.19(2.3)	12.79(3.01)	11.70(4.30)	9.27(2.01)	10.32(2.12)	11.42(3.90)
cSFGL	6.48(0.57)*	7.39(0.92)	8.35(0.78)	7.61(0.88)	7.95(1.21)	7.35(2.30)
Proposed	6.91(0.61)	7.04(0.88)	7.60(0.69)*	6.66(0.81)*	7.15(1.08)	6.87(1.85)*

is performed using the Analysis of variance (ANOVA) test to investigate the significant difference (P-value ≤ 0.05) of results at each time point in terms of RMSE which are indicated in Table 5.6. This table shows that our proposed method significantly outperforms the other well-established methods of lasso, ridge, and FGL for all time points and TGL and cFSGL at most of the time points for predicting MMSE and ADAS.

5.4 Discussion

In this study, the prediction of 4 important cognitive scores have been evaluated using multimodal neuroimaging with the focus placed on identifying CN and MCI individuals who later converted to MCI or to AD. We considered 5 groups of converter/stable normal cognition/mild cognitive impairment and AD with a time window of 4 years including 6

time points. A 10-fold cross-validation was used to prevent any bias in the training and testing sets and an ensemble approach was applied to the primary predictions of separate-modality regressions to improve the overall estimation of prediction. We proposed a kernel-based model to capture the nonlinear associations between the separate modalities of neuroimaging and neuropsychological test scores.

We investigated the most relevant MRI features that have the highest weights in the training phase to predict the cognitive scores. We found that white matter volume of left hippocampus/ left inferior parietal/ right inferior lateral ventricle, cortical thickness average of left middle temporal/ right entorhinal/ right inferior parietal/ left temporal pole, the cortical volume of right pars opercularis/ left insula, and surface area of left supramarginal are among the stable features to predict MMSE, ADAS, and RAVLT tests. These observations agree with previous studies in the pathological pathway of Alzheimer's disease, which show that the medial temporal lobe, including the entorhinal and hippocampus cortices, is the first to be affected by disease progression [105, 138, 139]. Moreover, bilateral hippocampal and parahippocampal regions are found among the most important features in gauging disease progression [100, 140] as well as thickness average and volume of inferior parietal [105, 141]. Moradi et al. found the medial temporal lobe and amygdala as the top predictors to estimate RAVL-Im and hippocampus, angular gyrus, and amygdala as the top predictors for RAVLT-PF [128]. The medial temporal lobe especially the hippocampus plays an important role in episodic memory specifically for the establishment and keeping the memories before storing them to other areas [142, 143].

Although these findings agree with other investigations to find those stable features in disease progression, the large effect of MCI or Alzheimer's pathology may overshadow the cognitively normal group. Since all subjects are considered as a pool, the weight of stable

features is more significant when subjects already progressed to MCI and AD. Therefore, we studied the most relevant features in detecting CNs vs. CNc groups. We observed that cortical thickness average of right precuneus/ left inferior temporal/left insula, white matter volume of right choroid plexus/right thalamus/left ventral DC, the surface area of left rostral anterior cingulate/ right superior parietal/ left lateral orbitofrontal/ right pars triangularis/ left parahippocampal and cortical volume of right rostral middle frontal are among the most important features. Previous studies found structural changes in right precuneus, superior frontal, and left thalamus [144] as well as the hippocampus, entorhinal cortex, and ventricles [145, 146] as the most important MRI features for conversion from CN to MCI which are in line with our observations. We also observed that transition from CN to MCI or AD is more associated with parietal lobe and prefrontal cortex baseline atrophy [147–150]. Recent studies also suggest that the insular cortex which is associated with taste and non-taste recognition memory by interaction with the amygdala as part of the default mode network was discovered to be disrupted in AD [151, 152]. Our results demonstrate that in addition to the amygdala and well-known hippocampus, insula, and middle temporal are also related to verbal episodic memory as reflected in the prediction of the RVALT scores.

Previous works studied the predictive models for different neuroimaging techniques in Alzheimer's trajectory. Stonning et al. applied relevance vector regression on 586 subjects using MRI and CSF modalities and achieved RMSE/R² of 2.19/0.57 and 7.27/0.59 for prediction of MMSE and ADAS at baseline, respectively as well as 18.97/0.19 for predicting RAVLT-PF [153]. Zhang and Shen obtained R² of 0.51 for the prediction of MMSE and 0.53 for ADAS changes in a 2-year study with a total number of 186 subjects and using a multimodal multitask approach including the MRI, PET, and CSF [79]. Zhang et al. achieved RMSE of 2.61 and 2.52 with a correlation of 0.73 for prediction of MMSE

as well as RMSE of 5.18 and 5.27 with the correlation of 0.77 for prediction of ADAS at month 12 and 48, respectively [63]. Moradi et al. used elastic net linear regression for predicting the two tasks of RAVLT-Im and RAVLT-PF from MRI measurements and obtained normalized RMSE (nRMSE) of 0.87 along with correlation of 0.5 and nRMSE of 0.9 along with a correlation of 0.43 for the prediction of RAVLT-Im and RAVLT-PF scores, respectively [128]. Jie et al. used a temporally-constrained group sparse lasso using MRI measurements with the time window of 24 months and achieved RMSE/R² of 2.84/0.65 for predicting MMSE and 5.85/0.67 for predicting ADAS [89]. Liu et al. proposed a multitask sparse group lasso learning and obtained RMSE/R² of 2.16/0.52 and 6.59/0.66 for predicting MMSE and ADAS, respectively using the baseline MRI [154]. In another study the correlation of 0.66 and 0.7 using $l_{2,1}$, l_1 regularized multitask regression applied on longitudinal data (3-year time window) of MRI, PET, APOE, age, and education were obtained for MMSE and ADAS, respectively [92].

It is worth noting that most of the aforementioned studies excluded the CNc subjects in their dataset, while we included the CNc group in order to identify individuals at risk for progressing into MCI or AD. There are few studies focusing on normal cognition conversion. Some studies investigated the conversion from CN to MCI and AD to test for the inflection points before the diagnosis of AD [155], to investigate the probable presymptomatic markers in healthy aging [156], and to estimate if the age of symptom onset could be similar across generation in subjects with Alzheimer’s history in at least one parent [157]. Zhan et al. considered a training set of 120 CN and 121 AD subjects, as well as a test set of 20 CNs and 20 CNc subjects citezhan2015identification. They achieved 70% accuracy for classification of CNs vs. CNc using multimodal neuroimaging of MRI, FDG-PET, and AV45-PET in the follow-up for 24 months.

CHAPTER 6

CONCLUSION

This study aimed to develop several machine learning algorithms to diagnose the Alzheimer's disease in its early stage, preferably in its presymptomatic phase where no manifestation of cognitive decline is yet apparent. To fulfill this aim, two types of cross-sectional and longitudinal data analyses have been performed for classifying the different subject groups as well as for predicting their future cognitive status using the most common cognitive tests.

In the second chapter, a machine learning approach based on a deep neural network has been proposed for binary and multiclass classifications with a focus placed on the delineation of EMCI group from the CN controls. The proposed approach introduces the use of the Adam algorithm to update the learning weights which improves the accuracy of diagnosis while efficiently updating the learning weights and decrease the time of convergence. We combined multimodal imaging of MRI and PET with the neuropsychological test scores through the well-known RAVLT, MoCA, and ECogT tests. The high accuracy of 84.0% for delineating the EMCI group from the CN group and accuracy of 96.8% for CN and AD classification were achieved. In addition, the proposed deep learning algorithm is successfully used for multiclass of CN, EMCI, LMCI, and AD classification as well. During the feature selection and training processes, we found that RAVLT and ECogT are useful neuropsychological tools for the early detection of Alzheimer's disease.

In the third chapter, an SVM-based approach with the RBF kernel has been proposed for binary classification with a focus placed on the delineation of the EMCI group from the cognitively normal group. Diagnosis of this early prodromal stage of AD could result in the planning of early treatment and therapeutic interventions to slow down the progression

of the disease. We combined multimodal imaging of MRI and PET with the neuropsychological test scores through the well-known RAVLT, MoCA, and ECogT tests. The high accuracy of 81.1% for delineating the EMCI group from the CN group and accuracy of 96.2% for CN and AD classification were achieved. In addition, the accuracy of 75.6% for CN and EMCI classification has been obtained using only MRI, amyloid PET, age, and education without using neuropsychological test scores which is higher than all of the results obtained from previous studies.

In chapter four, a probabilistic approach for finding the most important features augmented with a Gaussian-based model is designed to address the challenging classification of the EMCI group from the CN group. This approach evaluated the merits of using the Gaussian process and integrating the Bayesian prediction and classification model as another direction for the application of machine learning in AD. To this end, we have used a feature selection method based on the random forest algorithm and applied a recursive feature elimination (RF-RFE) approach. Many of the related studies involve the MCI group as a whole in their classification without distinction of the early and late stages of MCI (EMCI and LMCI), which takes away the ability to detect the earliest signs of AD, a challenge which this study has aimed to resolve. In addition, a large number of subjects was considered, proving the ability of the proposed method to be generalized for other classification purposes.

In chapter five, we introduced a multitask framework to model the disease progression on a longitudinal data considering the structured sparsity of features with both coupling across tasks and group selection for individual tasks. This framework is optimized by the ADMM algorithm to solve for the non-smooth objective formulation. The proposed framework includes a kernel-based approach to capture the nonlinear relationship between

neuroimaging data and the cognitive scores along with a decision-tree ensemble to fuse the separate modalities seeking a strong prediction performance. Our proposed method provides a unified approach to fuse heterogeneous data that cannot be directly concatenated due to temporal sparsity of modalities as well as allowing for more flexibility in having different weights for different modalities by specific-regression training. One advantage of this study relates to the indication of at risk groups of normal cognition and mild cognitive impairment to delay the onset of symptoms at clinical trials.

In addition, the most important biomarkers can be identified due to the sparsity-inducing nature of the algorithm. Four cognitive scores of MMSE, ADAS, RAVLT-Im, and RAVLT-PF were considered for predictive purposes using the proposed multimodal neuroimaging platform. The episodic tests as used in this study were shown to be effective for selecting the at-risk groups. We observed that MRI morphometry was the most sensitive biomarker to predict the conversion and we realized that parietal and prefrontal cortices are also associated with episodic memory in addition to the temporal lobe. Although adding FDG-PET, CSF, and APOE allele improved the prediction error at all 6 time points and significantly at 4 time points, we observed that using multimodal neuroimaging does not statistically enhance the prediction performance at some time points due to the large size of missing data. It is clear that for longitudinal studies of this type (4-year duration), the missing data challenge remains the most difficult to overcome for the development of any future machine learning algorithm for predicting disease progression.

Limitations

Given the many accomplishments made through this research endeavor, there are still some limitations to this study that need to be overcome. First, there are obviously other

biomarkers such as functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) that may potentially augment and improve these current classification and prediction results, especially in terms of the CN vs. EMCI classification and CN conversion groups. In addition, the presence of noise has not been tested in our study since the MRI and PET images considered underwent a high level of quality control. It is essential for any study involving the acquisition of any imaging modality to undergo quality control to ensure that no shading or aliasing is experienced and that such images are free from the presence of impulse noise, Gaussian noise, and any other source of noise that could affect the results of segmentation and the eventual extraction of key structural or functional features. Moreover, in chapter 4 since finding the optimal threshold for the Pearson's correlation using optimization algorithms is time-consuming, we set the threshold manually in this study. However, we contemplate to investigate using the mean (μ), standard deviation (σ) and the upper percentile (z) to determine a practical threshold of what we assume will be normal distributions of the data under study as $T = \mu + z\sigma$, a formula we discovered to work well with electroencephalography (EEG) data in epilepsy.

Furthermore, in chapter five, although β -amyloid of CSF correlates with the amyloid level of PET neuroimaging, the two are not identical. Therefore, using the amyloid level measured by PET as well as the CSF p-tau level using tau imaging instead of CSF may change the results. In addition, the age-correction procedure could improve the learning process to predict the cognitive scores, which was not added in our model. Also, one challenge in clinical data processing is missing values. Although we handled the problem of missing data using a decision tree-based model to fuse the separate modalities, still the observations that had missing values in the target were excluded in this study.

Therefore, these limitations will guide the development of new machine learning algorithms that would improve these results and to do multiclass classification using a more comprehensive multimodal neuroimaging platform. In addition, using unsupervised approaches could exploit the information of unlabeled data for classifying and predicting conversions in the presence of missing data.

REFERENCES

- [1] Alzheimer's Association et al. 2018 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 14(3):367–429, 2018.
- [2] Ronald C Petersen, Rosebud O Roberts, David S Knopman, Bradley F Boeve, Yonas E Geda, Robert J Ivnik, Glenn E Smith, and Clifford R Jack. Mild cognitive impairment: ten years later. *Archives of neurology*, 66(12):1447–1455, 2009.
- [3] Eric Westman, Andrew Simmons, Yi Zhang, J-Sebastian Muehlboeck, Catherine Tunnard, Yawu Liu, Louis Collins, Alan Evans, Patrizia Mecocci, Bruno Vellas, et al. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. *Neuroimage*, 54(2):1178–1187, 2011.
- [4] Sidong Liu, Weidong Cai, Lingfeng Wen, and Dagan Feng. Multi-channel brain atrophy pattern analysis in neuroimaging retrieval. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pages 202–205. IEEE, 2013.
- [5] Norman L Foster, Judith L Heidebrink, Christopher M Clark, William J Jagust, Steven E Arnold, Nancy R Barbas, Charles S DeCarli, R Scott Turner, Robert A Koeppe, Roger Higdon, et al. Fdg-pet improves accuracy in distinguishing frontotemporal dementia and alzheimer's disease. *Brain*, 130(10):2616–2635, 2007.
- [6] Randall J Bateman, Chengjie Xiong, Tammie LS Benzinger, Anne M Fagan, Alison Goate, Nick C Fox, Daniel S Marcus, Nigel J Cairns, Xianyun Xie, Tyler M Blazey, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *New England Journal of Medicine*, 367(9):795–804, 2012.
- [7] Lindsay R Clark, Annie M Racine, Rebecca L Kosciak, Ozioma C Okonkwo, Corinne D Engelman, Cynthia M Carlsson, Sanjay Asthana, Barbara B Bendlin, Rick Chappell, Christopher R Nicholas, et al. Beta-amyloid and cognitive decline in late middle age: findings from the wisconsin registry for Alzheimer's prevention study. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 12(7):805–814, 2016.
- [8] Shibani Singh, Anant Srivastava, Liang Mi, Richard J Caselli, Kewei Chen, Dhruvan Goradia, Eric M Reiman, and Yalin Wang. Deep-learning-based classification of fdg-pet data for Alzheimer's disease categories. In *13th International Conference on Medical Information Processing and Analysis*, volume 10572, page 105720J. International Society for Optics and Photonics, 2017.

- [9] Parisa Forouzaneshad, Alireza Abbaspour, Chen Fang, Mercedes Cabrerizo, David Loewenstein, Ranjan Duara, and Malek Adjouadi. A survey on applications and analysis methods of functional magnetic resonance imaging for Alzheimer’s disease. *Journal of neuroscience methods*, 2018.
- [10] Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrero, Ben Glocker, and Daniel Rueckert. Disease prediction using graph convolutional networks: Application to autism spectrum disorder and alzheimer’s disease. *Medical image analysis*, 48:117–130, 2018.
- [11] Xiaowei Zhang, Bin Hu, Xu Ma, and Linxin Xu. Resting-state whole-brain functional connectivity networks for MCI classification using l2-regularized logistic regression. *IEEE transactions on nanobioscience*, 14(2):237–247, 2015.
- [12] Chong-Yaw Wee, Sen Yang, Pew-Thian Yap, Dinggang Shen, Alzheimer’s Disease Neuroimaging Initiative, et al. Sparse temporally dynamic resting-state functional connectivity networks for early mci identification. *Brain imaging and behavior*, 10(2):342–356, 2016.
- [13] Chunfei Li, Chen Fang, Malek Adjouadi, Mercedes Cabrerizo, Armando Barreto, Jean Andrian, Ranjan Duara, and David Loewenstein. A neuroimaging feature extraction model for imaging genetics with application to Alzheimer’s disease. In *Bioinformatics and Bioengineering (BIBE), 2017 IEEE 17th International Conference on*, pages 15–20. IEEE, 2017.
- [14] Chunfei Li, David A Loewenstein, Ranjan Duara, Mercedes Cabrerizo, Warren Barker, and Malek Adjouadi. The relationship of brain amyloid load and apoe status to regional cortical thinning and cognition in the adni cohort. *Journal of Alzheimer’s Disease*, 59(4):1269–1282, 2017.
- [15] Stefan Klöppel, Cynthia M Stonnington, Carlton Chu, Bogdan Draganski, Rachael I Scahill, Jonathan D Rohrer, Nick C Fox, Clifford R Jack Jr, John Ashburner, and Richard SJ Frackowiak. Automatic classification of MR scans in Alzheimer’s disease. *Brain*, 131(3):681–689, 2008.
- [16] Omid Kohannim, Xue Hua, Derrek P Hibar, Suh Lee, Yi-Yu Chou, Arthur W Toga, Clifford R Jack, Michael W Weiner, and Paul M Thompson. Boosting power for clinical trials using classifiers based on multiple biomarkers. *Neurobiology of aging*, 31(8):1429–1442, 2010.
- [17] Siqi Liu, Sidong Liu, Weidong Cai, Hangyu Che, Sonia Pujol, Ron Kikinis, Dagan Feng, Michael J Fulham, et al. Multimodal neuroimaging feature learning for

- multiclass diagnosis of Alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, 62(4):1132–1140, 2015.
- [18] Eric Westman, J-Sebastian Muehlboeck, and Andrew Simmons. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *Neuroimage*, 62(1):229–238, 2012.
- [19] Lei Yuan, Yalin Wang, Paul M Thompson, Vaibhav A Narayan, Jieping Ye, Alzheimer's Disease Neuroimaging Initiative, et al. Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. *NeuroImage*, 61(3):622–632, 2012.
- [20] Vamsi K Ithapu, Vikas Singh, Ozioma C Okonkwo, Richard J Chappell, N Maritza Dowling, and Sterling C Johnson. Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 11(12):1489–1499, 2015.
- [21] Ranjan Duara, David A Loewenstein, Qian Shen, Warren Barker, Elizabeth Potter, Daniel Varon, Kristen Heurlin, Rik Vandenberghe, and Christopher Buckley. Amyloid positron emission tomography with 18F-flutemetamol and structural magnetic resonance imaging in the classification of mild cognitive impairment and Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 9(3):295–301, 2013.
- [22] Patricio Andres Donnelly-Kehoe, Guido Orlando Pascariello, Juan Carlos Gómez, Alzheimers Disease Neuroimaging Initiative, et al. Looking for Alzheimer's disease morphometric signatures using machine learning techniques. *Journal of neuroscience methods*, 302:24–34, 2018.
- [23] Mingxia Liu, Daoqiang Zhang, and Dinggang Shen. Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment. *IEEE transactions on medical imaging*, 35(6):1463–1474, 2016.
- [24] Jieping Ye, Teresa Wu, Jing Li, and Kewei Chen. Machine learning approaches for the neuroimaging study of Alzheimer's disease. *Computer*, 44(4):99–101, 2011.
- [25] Fayao Liu, Luping Zhou, Chunhua Shen, and Jianping Yin. Multiple kernel learning in the primal for multimodal Alzheimer's disease classification. *IEEE journal of biomedical and health informatics*, 18(3):984–990, 2014.

- [26] Christos Davatzikos, Priyanka Bhatt, Leslie M Shaw, Kayhan N Batmanghelich, and John Q Trojanowski. Prediction of mci to ad conversion, via mri, csf biomarkers, and pattern classification. *Neurobiology of aging*, 32(12):2322–e19, 2011.
- [27] Simon F Eskildsen, Pierrick Coupé, Daniel García-Lorenzo, Vladimir Fonov, Jens C Pruessner, D Louis Collins, Alzheimer’s Disease Neuroimaging Initiative, et al. Prediction of alzheimer’s disease in subjects with mild cognitive impairment from the adni cohort using patterns of cortical thinning. *Neuroimage*, 65:511–521, 2013.
- [28] Federico De Martino, Giancarlo Valente, Noël Staeren, John Ashburner, Rainer Goebel, and Elia Formisano. Combining multivariate voxel selection and support vector machines for mapping and classification of fmri spatial patterns. *Neuroimage*, 43(1):44–58, 2008.
- [29] Parisa Forouzaneshad, Alireza Abbaspour, Mercedes Cabrerizo, and Malek Adioudi. Early diagnosis of mild cognitive impairment using random forest feature selection. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4. IEEE, 2018.
- [30] Juergen Dukart, Karsten Mueller, Henryk Barthel, Arno Villringer, Osama Sabri, and Matthias Leopold Schroeter. Meta-analysis based svm classification enables accurate detection of alzheimer’s disease across different clinical centers using fdg-pet and mri. *Psychiatry Research: Neuroimaging*, 212(3):230–236, 2013.
- [31] Malek Adjouadi, Nuannuan Zong, and Melvin Ayala. Multidimensional pattern recognition and classification of white blood cells using support vector machines. *Particle & Particle Systems Characterization*, 22(2):107–118, 2005.
- [32] Talia M Nir, Julio E Villalon-Reina, Gautam Prasad, Neda Jahanshad, Shantanu H Joshi, Arthur W Toga, Matt A Bernstein, Clifford R Jack, Michael W Weiner, and Paul M Thompson. Diffusion weighted imaging-based maximum density path analysis and classification of Alzheimer’s disease. *Neurobiology of aging*, 36:S132–S140, 2015.
- [33] M López, J Ramírez, JM Górriz, D Salas-Gonzalez, I Alvarez, F Segovia, and CG Puntonet. Automatic tool for Alzheimer’s disease diagnosis using PCA and bayesian classification rules. *Electronics Letters*, 45(8):389–391, 2009.
- [34] Sidong Liu, Yang Song, Weidong Cai, Sonia Pujol, Ron Kikinis, Xiaogang Wang, and Dagan Feng. Multifold bayesian kernelization in Alzheimer’s diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 303–310. Springer, 2013.

- [35] Flávio Luiz Seixas, Bianca Zadrozny, Jerson Laks, Aura Conci, and Débora Christina Muchaluat Saade. A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer’s disease and mild cognitive impairment. *Computers in biology and medicine*, 51:140–158, 2014.
- [36] Claudia Plant, Stefan J Teipel, Annahita Oswald, Christian Böhm, Thomas Meindl, Janaina Mourao-Miranda, Arun W Bokde, Harald Hampel, and Michael Ewers. Automated detection of brain atrophy patterns based on mri for the prediction of alzheimer’s disease. *Neuroimage*, 50(1):162–174, 2010.
- [37] Dongren Yao, Vince D Calhoun, Zening Fu, Yuhui Du, and Jing Sui. An ensemble learning system for a 4-way classification of Alzheimer’s disease and mild cognitive impairment. *Journal of neuroscience methods*, 302:75–81, 2018.
- [38] Tom Brosch, Roger Tam, Alzheimer’s Disease Neuroimaging Initiative, et al. Manifold learning of brain mris by deep learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 633–640. Springer, 2013.
- [39] Siqi Liu, Sidong Liu, Weidong Cai, Sonia Pujol, Ron Kikinis, and Dagan Feng. Early diagnosis of Alzheimer’s disease with deep learning. In *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, pages 1015–1018. IEEE, 2014.
- [40] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer’s Disease Neuroimaging Initiative, et al. Latent feature representation with stacked auto-encoder for ad/mci diagnosis. *Brain Structure and Function*, 220(2):841–859, 2015.
- [41] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.
- [42] Nicola Amoroso, Domenico Diacono, Annarita Fanizzi, Marianna La Rocca, Alfonso Monaco, Angela Lombardi, Cataldo Guaragnella, Roberto Bellotti, Sabina Tangaro, Alzheimer’s Disease Neuroimaging Initiative, et al. Deep learning reveals Alzheimer’s disease onset in MCI subjects: Results from an international challenge. *Journal of neuroscience methods*, 302:3–9, 2018.
- [43] Parisa Forouzaneshad, Alireza Abbaspour, Chunfei Li, Mercedes Cabrerizo, and Malek Adjouadi. A deep neural network approach for early diagnosis of mild cognitive impairment using multiple features. In *2018 17th IEEE International*

Conference on Machine Learning and Applications (ICMLA), pages 1341–1346. IEEE, 2018.

- [44] Yue Cui, Wei Wen, Darren M Lipnicki, Mirza Faisal Beg, Jesse S Jin, Suhuai Luo, Wanlin Zhu, Nicole A Kochan, Simone Reppermund, Lin Zhuang, et al. Automated detection of amnesic mild cognitive impairment in community-dwelling elderly adults: a combined spatial atrophy and white matter alteration approach. *Neuroimage*, 59(2):1209–1217, 2012.
- [45] Andres Ortiz, Jorge Munilla, Juan M Gorriz, and Javier Ramirez. Ensembles of deep learning architectures for the early diagnosis of the alzheimer’s disease. *International journal of neural systems*, 26(07):1650025, 2016.
- [46] Chen Fang, Chunfei Li, Mercedes Cabrerizo, Armando Barreto, Jean Andrian, Naphtali Rishe, David Loewenstein, Ranjan Duara, and Malek Adjouadi. Gaussian discriminant analysis for optimal delineation of mild cognitive impairment in Alzheimer’s disease. *International Journal of Neural Systems*, 2018.
- [47] Edward Challis, Peter Hurley, Laura Serra, Marco Bozzali, Seb Oliver, and Mara Cercignani. Gaussian process classification of Alzheimer’s disease and mild cognitive impairment from resting-state fMRI. *Neuroimage*, 112:232–243, 2015.
- [48] Reisa A Sperling, Paul S Aisen, Laurel A Beckett, David A Bennett, Suzanne Craft, Anne M Fagan, Takeshi Iwatsubo, Clifford R Jack, Jeffrey Kaye, Thomas J Montine, et al. Toward defining the preclinical stages of Alzheimer’s disease: Recommendations from the national institute on Aging-Alzheimer’s association workgroups on diagnostic guidelines for alzheimer’s disease. *Alzheimer’s & dementia: the journal of the Alzheimer’s Association*, 7(3):280–292, 2011.
- [49] Randy L Buckner. Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. *Neuron*, 44(1):195–208, 2004.
- [50] Reinhold Schmidt, Helena Schmidt, J David Curb, Kamal Masaki, Lon R White, and Lenore J Launer. Early inflammation and dementia: A 25-year follow-up of the honolulu-asia aging study. *Annals of neurology*, 52(2):168–174, 2002.
- [51] Jeffrey L Cummings, Harry V Vinters, Gregory M Cole, and Zaven S Khachaturian. Alzheimer’s disease etiologies, pathophysiology, cognitive reserve, and treatment opportunities. *Neurology*, 51(1 Suppl 1):S2–S17, 1998.

- [52] Ranjan Duara, David A Loewenstein, Elizabeth Potter, Warren Barker, Ashok Raj, Michael Schoenberg, Yougui Wu, Jessica Banko, Huntington Potter, Maria T Greig, et al. Pre-MCI and MCI: neuropsychological, clinical, and imaging features and progression rates. *The American Journal of Geriatric Psychiatry*, 19(11):951–960, 2011.
- [53] Gautam Prasad, Shantanu H Joshi, Talia M Nir, Arthur W Toga, and Paul M Thompson. Brain connectivity and novel network measures for Alzheimer’s disease classification. *Neurobiology of aging*, 36:S121–S131, 2015.
- [54] Mohammed Goryawala, Qi Zhou, Warren Barker, David A Loewenstein, Ranjan Duara, and Malek Adjouadi. Inclusion of neuropsychological scores in atrophy models improves diagnostic classification of alzheimer’s disease and mild cognitive impairment. *Computational intelligence and neuroscience*, 2015:56, 2015.
- [55] Hiroyuki Shimada, Hyuma Makizako, Daisuke Yoshida, Hiroshi Shimokata, Kengo Ito, Yukihiro Washimi, Hidetoshi Endo, Takao Suzuki, et al. Characteristics of cognitive function in early and late stages of amnesic mild cognitive impairment. *Geriatrics & gerontology international*, 13(1):83–89, 2013.
- [56] Ricardo Guerrero, Robin Wolz, AW Rao, Daniel Rueckert, Alzheimer’s Disease Neuroimaging Initiative (ADNI, et al. Manifold population modeling as a neuroimaging biomarker: application to ADNI and ADNI-GO. *NeuroImage*, 94:275–286, 2014.
- [57] Shashank Tripathi, Seyed Hossein Nozadi, Mahsa Shakeri, and Samuel Kadoury. Sub-cortical shape morphology and voxel-based features for Alzheimer’s disease classification. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pages 991–994. IEEE, 2017.
- [58] Hao Guo, Fan Zhang, Junjie Chen, Yong Xu, and Jie Xiang. Machine learning classification combining multiple features of a hyper-network of fMRI data in Alzheimer’s Disease. *Frontiers in neuroscience*, 11:615, 2017.
- [59] Mahsa Shakeri, Herve Lombaert, Shashank Tripathi, Samuel Kadoury, Alzheimer’s Disease Neuroimaging Initiative, et al. Deep spectral-based shape features for Alzheimer’s disease classification. In *International Workshop on Spectral and Shape Analysis in Medical Imaging*, pages 15–24. Springer, 2016.
- [60] Qi Zhou, Mohammed Goryawala, Mercedes Cabrerizo, Jin Wang, Warren Barker, David A Loewenstein, Ranjan Duara, and Malek Adjouadi. An optimal decisional

space for the classification of Alzheimer's disease and mild cognitive impairment. *IEEE Transactions on Biomedical Engineering*, 61(8):2245–2253, 2014.

- [61] Biao Jie, Mingxia Liu, and Dinggang Shen. Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease. *Medical Image Analysis*, 47:81–94, 2018.
- [62] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Medical image analysis*, 37:101–113, 2017.
- [63] Jie Zhang, Qingyang Li, Richard J Caselli, Paul M Thompson, Jieping Ye, and Yalin Wang. Multi-source multi-target dictionary learning for prediction of cognitive decline. In *International Conference on Information Processing in Medical Imaging*, pages 184–197. Springer, 2017.
- [64] Clifford R Jack Jr, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.
- [65] Clifford R Jack Jr, Heather J Wiste, Christopher G Schwarz, Val J Lowe, Matthew L Senjem, Prashanthi Vemuri, Stephen D Weigand, Terry M Therneau, Dave S Knopman, Jeffrey L Gunter, et al. Longitudinal tau pet in ageing and alzheimer's disease. *Brain*, 141(5):1517–1528, 2018.
- [66] Paul M Thompson and Arthur W Toga. Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations. *Medical image analysis*, 1(4):271–294, 1997.
- [67] Liana G Apostolova, Lisa Mosconi, Paul M Thompson, Amity E Green, Kristy S Hwang, Anthony Ramirez, Rachel Mistur, Wai H Tsui, and Mony J de Leon. Subregional hippocampal atrophy predicts alzheimer's dementia in the cognitively normal. *Neurobiology of aging*, 31(7):1077–1088, 2010.
- [68] Simon Duchesne, Anna Caroli, Cristina Geroldi, D Louis Collins, and Giovanni B Frisoni. Relating one-year cognitive change in mild cognitive impairment to baseline mri features. *Neuroimage*, 47(4):1363–1370, 2009.
- [69] Rosie E Curiel, David A Loewenstein, Monica Rosselli, Ailyn Penate, Maria T Greig-Custo, Russell M Bauer, Salvador M Guinjoan, Kevin S Hanson, Chunfei Li,

Gabriel Lizarraga, et al. Semantic intrusions and failure to recover from semantic interference in mild cognitive impairment: relationship to amyloid and cortical thickness. *Current Alzheimer Research*, 15(9):848–855, 2018.

- [70] Ranjan Duara, Warren Barker, David Loewenstein, Maria T Greig, Rosemarie Rodriguez, Mohammed Goryawala, Qi Zhou, and Malek Adjouadi. Insights into cognitive aging and Alzheimer’s disease using amyloid PET and structural MRI scans. *Clinical and Translational Imaging*, 3(1):65–74, 2015.
- [71] Gerard Martí-Juan, Gerard Sanroma-Guell, and Gemma Piella. A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in alzheimer’s disease. *Computer Methods and Programs in Biomedicine*, 189:105348, 2020.
- [72] Katja Franke and Christian Gaser. Longitudinal changes in individual brainage in healthy aging, mild cognitive impairment, and alzheimer’s disease. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry*, 25(4):235, 2012.
- [73] Christos Davatzikos, Feng Xu, Yang An, Yong Fan, and Susan M Resnick. Longitudinal progression of alzheimer’s-like patterns of atrophy in normal older adults: the spare-ad index. *Brain*, 132(8):2026–2035, 2009.
- [74] Ricardo Guerrero, Alexander Schmidt-Richberg, Christian Ledig, Tong Tong, Robin Wolz, Daniel Rueckert, Alzheimer’s Disease Neuroimaging Initiative (ADNI, et al. Instantiated mixed effects modeling of alzheimer’s disease markers. *NeuroImage*, 142:113–125, 2016.
- [75] Bruno M Jernak, Andrew Lang, Bo Liu, Elyse Katz, Yanwei Zhang, Bradley T Wyman, David Raunig, C Pierre Jernak, Brian Caffo, Jerry L Prince, et al. A computational neurodegenerative disease progression score: method and results with the alzheimer’s disease neuroimaging initiative cohort. *Neuroimage*, 63(3):1478–1486, 2012.
- [76] Ramon Casanova, Ryan T Barnard, Sarah A Gaussoin, Santiago Saldana, Kathleen M Hayden, JoAnn E Manson, Robert B Wallace, Stephen R Rapp, Susan M Resnick, Mark A Espeland, et al. Using high-dimensional machine learning methods to estimate an anatomical risk factor for alzheimer’s disease across imaging databases. *NeuroImage*, 183:401–411, 2018.
- [77] Vanessa H Clark, Susan M Resnick, Jimit Doshi, Lori L Beason-Held, Yun Zhou, Luigi Ferrucci, Dean F Wong, Michael A Kraut, and Christos Davatzikos. Longitudinal imaging pattern analysis (spare-cd index) detects early structural and

functional changes before cognitive decline in healthy older adults. *Neurobiology of aging*, 33(12):2733–2745, 2012.

- [78] Alexander Schmidt-Richberg, Ricardo Guerrero, Christian Ledig, Helena Molina-Abril, Alejandro F Frangi, Daniel Rueckert, Alzheimers Disease Neuroimaging Initiative, et al. Multi-stage biomarker models for progression estimation in alzheimer’s disease. In *International Conference on Information Processing in Medical Imaging*, pages 387–398. Springer, 2015.
- [79] Daoqiang Zhang, Dinggang Shen, Alzheimer’s Disease Neuroimaging Initiative, et al. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer’s disease. *NeuroImage*, 59(2):895–907, 2012.
- [80] Hua Wang, Feiping Nie, Heng Huang, Shannon Risacher, Chris Ding, Andrew J Saykin, and Li Shen. Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance. In *2011 International Conference on Computer Vision*, pages 557–562. IEEE, 2011.
- [81] Hua Wang, Feiping Nie, Heng Huang, Sungeun Kim, Kwangsik Nho, Shannon L Risacher, Andrew J Saykin, Li Shen, and Alzheimer’s Disease Neuroimaging Initiative. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the adni cohort. *Bioinformatics*, 28(2):229–237, 2012.
- [82] Daoqiang Zhang, Yaping Wang, Luping Zhou, Hong Yuan, Dinggang Shen, Alzheimer’s Disease Neuroimaging Initiative, et al. Multimodal classification of alzheimer’s disease and mild cognitive impairment. *Neuroimage*, 55(3):856–867, 2011.
- [83] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. Convex multi-task feature learning. *Machine learning*, 73(3):243–272, 2008.
- [84] Rie Kubota Ando and Tong Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6(Nov):1817–1853, 2005.
- [85] Jianhui Chen, Lei Tang, Jun Liu, and Jieping Ye. A convex formulation for learning shared structures from multiple tasks. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 137–144, 2009.

- [86] Laurent Jacob, Jean-philippe Vert, and Francis R Bach. Clustered multi-task learning: A convex formulation. In *Advances in neural information processing systems*, pages 745–752, 2009.
- [87] Jiayu Zhou, Jianhui Chen, and Jieping Ye. Clustered multi-task learning via alternating structure optimization. In *Advances in neural information processing systems*, pages 702–710, 2011.
- [88] Mingliang Wang, Daoqiang Zhang, Dinggang Shen, and Mingxia Liu. Multi-task exclusive relationship learning for alzheimer’s disease progression prediction with longitudinal data. *Medical image analysis*, 53:111–122, 2019.
- [89] Biao Jie, Mingxia Liu, Jun Liu, Daoqiang Zhang, and Dinggang Shen. Temporally constrained group sparse learning for longitudinal data analysis in alzheimer’s disease. *IEEE Transactions on Biomedical Engineering*, 64(1):238–249, 2016.
- [90] Xiaoli Liu, Peng Cao, Jianzhong Wang, Jun Kong, and Dazhe Zhao. Fused group lasso regularized multi-task feature learning and its application to the cognitive performance prediction of alzheimer’s disease. *Neuroinformatics*, 17(2):271–294, 2019.
- [91] Peng Cao, Xuanfeng Shan, Dazhe Zhao, Min Huang, and Osmar Zaiane. Sparse shared structure based multi-task learning for mri based cognitive performance prediction of alzheimer’s disease. *Pattern Recognition*, 72:219–235, 2017.
- [92] Peng Cao, Xiaoli Liu, Jinzhu Yang, Dazhe Zhao, Min Huang, and Osmar Zaiane. $l_{2,1-l_1}$ regularized nonlinear multi-task representation learning based cognitive performance prediction of alzheimer’s disease. *Pattern Recognition*, 79:195–215, 2018.
- [93] Parisa Forouzaneshad, Alireza Abbaspour, Chunfei Li, Chen Fang, Ulyana Williams, Mercedes Cabrerizo, Armando Barreto, Jean Andrian, Naphtali Rishe, Rosie E Curiel, et al. A gaussian-based model for early detection of mild cognitive impairment using multimodal neuroimaging. *Journal of Neuroscience Methods*, 333:108544, 2020.
- [94] Tong Tong, Katherine Gray, Qinquan Gao, Liang Chen, Daniel Rueckert, Alzheimer’s Disease Neuroimaging Initiative, et al. Multi-modal classification of alzheimer’s disease using nonlinear graph fusion. *Pattern recognition*, 63:171–181, 2017.

- [95] Kerstin Ritter, Julia Schumacher, Martin Weygandt, Ralph Buchert, Carsten Allefeld, John-Dylan Haynes, Alzheimer’s Disease Neuroimaging Initiative, et al. Multimodal prediction of conversion to alzheimer’s disease based on incomplete biomarkers. *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(2):206–215, 2015.
- [96] Stefan Teipel, Alexander Drzezga, Michel J Grothe, Henryk Barthel, Gaël Chételat, Norbert Schuff, Pawel Skudlarski, Enrica Cavedo, Giovanni B Frisoni, Wolfgang Hoffmann, et al. Multimodal imaging in alzheimer’s disease: validity and usefulness for early detection. *The Lancet Neurology*, 14(10):1037–1053, 2015.
- [97] Kunal P Patel, David T Wymer, Vinay K Bhatia, Ranjan Duara, and Chetan D Rajadhyaksha. Multimodality imaging of dementia: Clinical importance and role of integrated anatomic and molecular imaging. *RadioGraphics*, 40(1):200–222, 2020.
- [98] Siqi Liu, Sidong Liu, Weidong Cai, Hangyu Che, Sonia Pujol, Ron Kikinis, Dagan Feng, Michael J Fulham, et al. Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer’s disease. *IEEE Transactions on Biomedical Engineering*, 62(4):1132–1140, 2014.
- [99] Chris Hinrichs, Vikas Singh, Guofan Xu, Sterling C Johnson, Alzheimers Disease Neuroimaging Initiative, et al. Predictive markers for ad in a multi-modality framework: an analysis of mci progression in the adni population. *Neuroimage*, 55(2):574–589, 2011.
- [100] Daoqiang Zhang, Dinggang Shen, Alzheimer’s Disease Neuroimaging Initiative, et al. Predicting future clinical changes of mci patients using longitudinal and multimodal biomarkers. *PloS one*, 7(3):e33182, 2012.
- [101] Tianle Chen, Donglin Zeng, and Yuanjia Wang. Multiple kernel learning with random effects for predicting longitudinal outcomes and data integration. *Biometrics*, 71(4):918–928, 2015.
- [102] Chen Fang, Chunfei Li, Parisa Forouzaneshad, Mercedes Cabrerizo, Rosie E Curiel, David Loewenstein, Ranjan Duara, Malek Adjouadi, Alzheimer’s Disease Neuroimaging Initiative, et al. Gaussian discriminative component analysis for early detection of alzheimer’s disease: A supervised dimensionality reduction algorithm. *Journal of Neuroscience Methods*, 344:108856, 2020.
- [103] Goo-Rak Kwon, Yubraj Gupta, and Ramesh Kumar Lama. Prediction and classification of alzheimer’s disease based on combined features from apolipoprotein-e

- genotype, cerebrospinal fluid, mr, and fdg-pet imaging biomarkers. *Frontiers in computational neuroscience*, 13:72, 2019.
- [104] Shuo Xiang, Lei Yuan, Wei Fan, Yalin Wang, Paul M Thompson, Jieping Ye, Alzheimer’s Disease Neuroimaging Initiative, et al. Bi-level multi-source learning for heterogeneous block-wise missing data. *NeuroImage*, 102:192–206, 2014.
- [105] Jiayu Zhou, Jun Liu, Vaibhav A Narayan, Jieping Ye, Alzheimer’s Disease Neuroimaging Initiative, et al. Modeling disease progression via multi-task learning. *NeuroImage*, 78:233–248, 2013.
- [106] Jingwen Yan, Taiyong Li, Hua Wang, Heng Huang, Jing Wan, Kwangsik Nho, Sungeun Kim, Shannon L Risacher, Andrew J Saykin, Li Shen, et al. Cortical surface biomarkers for predicting cognitive outcomes using group l2, 1 norm. *Neurobiology of aging*, 36:S185–S193, 2015.
- [107] Hongwu Wang, Yanqing Liu, Shoulian Wei, and Zijun Yan. Comparative seasonal variation and chemical composition of essential oils from the leaves and stems of schefflera heptaphylla using microwave-assisted and conventional hydrodistillation. *Industrial Crops and Products*, 36(1):229–237, 2012.
- [108] Liqiang Nie, Luming Zhang, Lei Meng, Xueming Song, Xiaojun Chang, and Xuelong Li. Modeling disease progression via multisource multitask learners: A case study with alzheimer’s disease. *IEEE transactions on neural networks and learning systems*, 28(7):1508–1519, 2016.
- [109] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.
- [110] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [111] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [112] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238, 2005.

- [113] Kenichi Ota, Naoya Oishi, Kengo Ito, Hidenao Fukuyama, SEAD-J Study Group, Alzheimer’s Disease Neuroimaging Initiative, et al. Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer’s disease. *Journal of neuroscience methods*, 256:168–183, 2015.
- [114] Benson Mwangi, Tian Siva Tian, and Jair C Soares. A review of feature reduction techniques in neuroimaging. *Neuroinformatics*, 12(2):229–244, 2014.
- [115] Carlton Chu, Ai-Ling Hsu, Kun-Hsien Chou, Peter Bandettini, ChingPo Lin, Alzheimer’s Disease Neuroimaging Initiative, et al. Does feature selection improve classification accuracy? impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage*, 60(1):59–70, 2012.
- [116] Yue Cui, Bing Liu, Suhuai Luo, Xiantong Zhen, Ming Fan, Tao Liu, Wanlin Zhu, Mira Park, Tianzi Jiang, Jesse S Jin, et al. Identification of conversion from mild cognitive impairment to Alzheimer’s disease using multivariate predictors. *PloS one*, 6(7):e21896, 2011.
- [117] Katherine R Gray, Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Daniel Rueckert, Alzheimer’s Disease Neuroimaging Initiative, et al. Random forest-based similarity measures for multi-modal classification of Alzheimer’s disease. *NeuroImage*, 65:167–175, 2013.
- [118] Bjoern H Menze, B Michael Kelm, Ralf Masuch, Uwe Himmelreich, Peter Bachert, Wolfgang Petrich, and Fred A Hamprecht. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 10(1):213, 2009.
- [119] Vladimir Vapnik. *Statistical learning theory*. 1998. Wiley, New York, 1998.
- [120] Vladimir Naumovich Vapnik and Samuel Kotz. *Estimation of dependences based on empirical data*, volume 40. Springer-Verlag New York, 1982.
- [121] Andrew Gordon Wilson. Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes. *University of Cambridge*, 2014.
- [122] M Bishop Christopher. *PATTERN RECOGNITION AND MACHINE LEARNING*. Springer-Verlag New York, 2016.

- [123] Christopher KI Williams and David Barber. Bayesian classification with gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1342–1351, 1998.
- [124] R Core Team. R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. 2016, 2017.
- [125] Ingrid Arevalo-Rodriguez, Nadja Smailagic, Marta Roqué i Figuls, Agustín Ciapponi, Erick Sanchez-Perez, Antri Giannakou, Olga L Pedraza, Xavier Bonfill Cosp, and Sarah Cullum. Mini-mental state examination (mmse) for the detection of alzheimer’s disease and other dementias in people with mild cognitive impairment (mci). *Cochrane Database of Systematic Reviews*, (3), 2015.
- [126] Kimberly R Chapman, Hanaan Bing-Canar, Michael L Alosco, Eric G Steinberg, Brett Martin, Christine Chaisson, Neil Kowall, Yorghos Tripodis, and Robert A Stern. Mini mental state examination and logical memory scores for entry into alzheimer’s disease trials. *Alzheimer’s research & therapy*, 8(1):9, 2016.
- [127] Jacqueline K Kueper, Mark Speechley, and Manuel Montero-Odasso. The alzheimer’s disease assessment scale–cognitive subscale (adas-cog): modifications and responsiveness in pre-dementia populations. a narrative review. *Journal of Alzheimer’s Disease*, 63(2):423–444, 2018.
- [128] Elaheh Moradi, Ilona Hallikainen, Tuomo Hänninen, Jussi Tohka, Alzheimer’s Disease Neuroimaging Initiative, et al. Rey’s auditory verbal learning test scores can be predicted from whole brain mri in alzheimer’s disease. *NeuroImage: Clinical*, 13:415–427, 2017.
- [129] Harald Hampel, Kaj Blennow, Leslie M Shaw, Yvonne C Hoessler, Henrik Zetterberg, and John Q Trojanowski. Total and phosphorylated tau protein as biological markers of alzheimer’s disease. *Experimental gerontology*, 45(1):30–40, 2010.
- [130] Cees Mulder, Nicolaas A Verwey, Wiesje M van der Flier, Femke H Bouwman, Astrid Kok, Evert J van Elk, Philip Scheltens, and Marinus A Blankenstein. Amyloid- β (1–42), total tau, and phosphorylated tau as cerebrospinal fluid biomarkers for the diagnosis of alzheimer disease. *Clinical chemistry*, 56(2):248–253, 2010.
- [131] Gail VW Johnson and William H Stoothoff. Tau phosphorylation in neuronal cell function and dysfunction. *Journal of cell science*, 117(24):5721–5729, 2004.

- [132] D Blacker, JL Haines, L Rodes, H Terwedow, RCP Go, LE Harrell, RT Perry, SS Bassett, G Chase, D Meyers, et al. Apoe-4 and age at onset of alzheimer's disease: the nimh genetics initiative. *Neurology*, 48(1):139–147, 1997.
- [133] Emmanuelle Genin, Didier Hannequin, David Wallon, Kristel Slegers, Mikko Hiltunen, Onofre Combarros, María Jesús Bullido, Sebastiaan Engelborghs, Peter De Deyn, Claudine Berr, et al. Apoe and alzheimer disease: a major gene with semi-dominant inheritance. *Molecular psychiatry*, 16(9):903–907, 2011.
- [134] Chinthaka Kuruwita. A bayesian approach for bandwidth selection in kernel density estimation with censored data. 2006.
- [135] Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005.
- [136] Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.
- [137] Robert M Freund, Paul Grigas, Rahul Mazumder, et al. A new perspective on boosting in linear regression via subgradient optimization and relatives. *The Annals of Statistics*, 45(6):2328–2364, 2017.
- [138] André Delacourte, Jean-Philippe David, Nicolas Sergeant, L Buee, A Wattez, P Vermersch, F Ghozali, C Fallet-Bianco, F Pasquier, F Lebert, et al. The biochemical pathway of neurofibrillary degeneration in aging and alzheimer's disease. *Neurology*, 52(6):1158–1158, 1999.
- [139] BC Dickerson, TR Stoub, RC Shah, RA Sperling, RJ Killiany, MS Albert, BT Hyman, Deborah Blacker, and L Detoleto-Morrell. Alzheimer-signature mri biomarker predicts ad dementia in cognitively normal adults. *Neurology*, 76(16):1395–1402, 2011.
- [140] Hua Wang, Feiping Nie, Heng Huang, Jingwen Yan, Sungeun Kim, Shannon Risacher, Andrew Saykin, and Li Shen. High-order multi-task feature learning to identify longitudinal phenotypic markers for alzheimer's disease progression prediction. In *Advances in neural information processing systems*, pages 1277–1285, 2012.
- [141] Marcus Likeman, Valerie M Anderson, John M Stevens, Adam D Waldman, Alison K Godbolt, Chris Frost, Martin N Rossor, and Nick C Fox. Visual assessment

of atrophy on magnetic resonance imaging in the diagnosis of pathologically confirmed young-onset dementias. *Archives of neurology*, 62(9):1410–1415, 2005.

- [142] Larry R Squire and John T Wixted. The cognitive neuroscience of human memory since hm. *Annual review of neuroscience*, 34:259–288, 2011.
- [143] Woorim Jeong, Chun Kee Chung, and June Sic Kim. Episodic memory in aspects of large-scale brain networks. *Frontiers in human neuroscience*, 9:454, 2015.
- [144] Lin Zhuang, Perminder S Sachdev, Julian N Trollor, Nicole A Kochan, Simone Reppermund, Henry Brodaty, and Wei Wen. Microstructural white matter changes in cognitively normal individuals at risk of amnesic mci. *Neurology*, 79(8):748–754, 2012.
- [145] Clifford R Jack, MM Shiung, SD Weigand, PC O’Brien, JL Gunter, Bradley F Boeve, David S Knopman, GE Smith, RJ Ivnik, Eric George Tangalos, et al. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnesic mci. *Neurology*, 65(8):1227–1231, 2005.
- [146] Marilyn Albert, Yuxin Zhu, Abhay Moghekar, Susumu Mori, Michael I Miller, Anja Soldan, Corinne Pettigrew, Ola Selnes, Shanshan Li, and Mei-Cheng Wang. Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years. *Brain*, 141(3):877–887, 2018.
- [147] Yue Cui, Perminder S Sachdev, Darren M Lipnicki, Jesse S Jin, Suhuai Luo, Wanlin Zhu, Nicole A Kochan, Simone Reppermund, Tao Liu, Julian N Trollor, et al. Predicting the development of mild cognitive impairment: a new use of pattern recognition. *Neuroimage*, 60(2):894–901, 2012.
- [148] Chong-Yaw Wee, Pew-Thian Yap, Wenbin Li, Kevin Denny, Jeffrey N Browndyke, Guy G Potter, Kathleen A Welsh-Bohmer, Lihong Wang, and Dinggang Shen. Enriched white matter connectivity networks for accurate identification of mci patients. *Neuroimage*, 54(3):1812–1822, 2011.
- [149] Jennifer L Whitwell, Ronald C Petersen, Selamawit Negash, Stephen D Weigand, Kejal Kantarci, Robert J Ivnik, David S Knopman, Bradley F Boeve, Glenn E Smith, and Clifford R Jack. Patterns of atrophy differ among specific subtypes of mild cognitive impairment. *Archives of neurology*, 64(8):1130–1138, 2007.
- [150] Sarah J Greene, Ronald J Killiany, Alzheimer’s Disease Neuroimaging Initiative, et al. Subregions of the inferior parietal lobule are affected in the progression to alzheimer’s disease. *Neurobiology of aging*, 31(8):1304–1311, 2010.

- [151] Federico Bermudez-Rattoni. The forgotten insular cortex: its role on recognition memory formation. *Neurobiology of learning and memory*, 109:207–216, 2014.
- [152] Parisa Forouzaneshad, Alireza Abbaspour, Chen Fang, Mercedes Cabrerizo, David Loewenstein, Ranjan Duara, and Malek Adjouadi. A survey on applications and analysis methods of functional magnetic resonance imaging for alzheimer’s disease. *Journal of neuroscience methods*, 317:121–140, 2019.
- [153] Cynthia M Stonnington, Carlton Chu, Stefan Klöppel, Clifford R Jack Jr, John Ashburner, Richard SJ Frackowiak, Alzheimer Disease Neuroimaging Initiative, et al. Predicting clinical scores from magnetic resonance scans in alzheimer’s disease. *Neuroimage*, 51(4):1405–1413, 2010.
- [154] Xiaoli Liu, Peng Cao, André R Gonçalves, Dazhe Zhao, and Arindam Banerjee. Modeling alzheimer’s disease progression with fused laplacian sparse group lasso. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(6):1–35, 2018.
- [155] Jacob W Vogel, Etienne Vachon-Presseau, Alexa Pichet Binette, Angela Tam, Pierre Orban, Renaud La Joie, Mélissa Savard, Cynthia Picard, Judes Poirier, Pierre Bellec, et al. Brain properties predict proximity to symptom onset in sporadic alzheimer’s disease. *Brain*, 141(6):1871–1883, 2018.
- [156] Diane B Howieson, Alison Dame, Richard Camicioli, Gary Sexton, Haydeh Payami, and Jeffrey A Kaye. Cognitive markers preceding alzheimer’s dementia in the healthy oldest old. *Journal of the American Geriatrics Society*, 45(5):584–589, 1997.
- [157] David K Johnson, Martha Storandt, John C Morris, and James E Galvin. Longitudinal study of the transition from healthy aging to alzheimer disease. *Archives of neurology*, 66(10):1254–1259, 2009.
- [158] Ye Zhan, Kewei Chen, Xia Wu, Daoqiang Zhang, Jiakai Zhang, Li Yao, and Xiaojuan Guo. Identification of conversion from normal elderly cognition to alzheimer’s disease using multimodal support vector machine. *Journal of Alzheimer’s Disease*, 47(4):1057–1067, 2015.

VITA

PARISA FOROUZANNEZHAD

EDUCATION

2017-2020	Ph.D., Electrical & Computer Engineering Florida International University (FIU) Miami, Florida
2009-2012	M.S., Electrical Engineering Azad University Tehran, Iran
2003-2007	B.S., Electrical Engineering University of Sadjad Mashhad, Iran

PUBLICATIONS AND PRESENTATIONS

P. Forouzannezhad, E. Zarafshan, U. Morar, M. Cabrerizo, R. Duara, D. Lowenstein, and M. Adjouadi, "Prediction of Conversion from Normal Cognition and Mild Cognitive Impairment to Alzheimer's Disease", *Medical Image Analysis*, (Submitted).

E. Zarafshan, H. Rajaei, P. Forouzannezhad, U. Morar, M. Cabrerizo, and M. Adjouadi, "Characterizing Focal and Generalized Epileptic Networks Using Interictal EEG", *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, (Accepted).

U. Morar, H. Martin, W. Izquierdo, P. Forouzannezhad, E. Zarafshan, R. E. Curiel, M. Roselli, D. Lowenstein, R. Duara, and M. Adjouadi, "A Deep-Learning Approach for the prediction of Mini-Mental State Examination Score in a Multimodal Longitudinal Study", *2020 International Conference on Computational Science and Computational Intelligence*, (Accepted).

E. Zarafshan, H. Rajaei, P. Forouzannezhad, U. Morar, M. Cabrerizo, and M. Adjouadi, "Functional Brain Connectivity from Interictal EEG in Epilepsy", *American Epilepsy Society*, (Accepted).

C. Fang, C. Li, P. Forouzannezhad, M. Cabrerizo, R. E Curiel, D. Loewenstein, R. Duara, and M. Adjouadi, "Gaussian discriminative component analysis for early detection of Alzheimer's disease: A supervised dimensionality reduction algorithm", *Journal of Neu-*

rosience Methods, 344 (2020): 108856.

P. Forouzannezhad, A. Abbaspour, C. Li, C. Fang, U. Williams, M. Cabrerizo, A. Barreto, J. Andrian, N. Rishe, R. E. Curiel, D. Lowenstein, R. Dura, and M. Adjouadi, "A Gaussian-based model for early detection of mild cognitive impairment using multimodal neuroimaging", *Journal of Neuroscience Methods*, 333 (2020): 108544.

A. Abbaspour, A. Sargolzaei, P. Forouzannezhad, K. Yen, and A. I. Sarwat. "Resilient Control Design for Load Frequency Control System under False Data Injection Attacks", *IEEE Transactions on Industrial Electronics*, 67, no. 9 (2019): 7951-7962.

A. Abbaspour, K. Yen, P. Forouzannezhad, A. Sargolzaei, "An Adaptive Resilient Control Approach for Pressure Control in Proton Exchange Membrane Fuel Cells", *IEEE Transactions on Industry Applications*, 55, no. 6 (2019): 6344-6354.

P. Forouzannezhad, A. Abbaspour, C. Fang, M. Cabrerizo, D. Lowenstein, R. Duara, and M. Adjouadi, " A Survey on Applications and Analysis Methods of Functional Magnetic Resonance Imaging for Alzheimer's Disease ", *Journal of Neuroscience Methods*, 317 (2019): 121-140.

P. Forouzannezhad, A. Abbaspour, C. Li, M. Cabrerizo, M. Adjouadi, "A Deep Neural Network Approach for Early Diagnosis of Mild Cognitive Impairment Using Multiple Features", *IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1341-1346. IEEE, 2018.

P. Forouzannezhad, A. Abbaspour, M. Cabrerizo, M. Adjouadi, "Early Diagnosis of Mild Cognitive Impairment Using Random Forest Feature Selection", *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1-4, 2018.

A. Abbaspour, K. Yen, P. Forouzannezhad, A. Sargolzaei, "A Neural Adaptive Approach for Active Fault-Tolerant Control Design in UAV", *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, (2018).

A. Abbaspour, K. Yen, P. Forouzannezhad, A. Sargolzaei, "Active Adaptive Fault-Tolerant Control Design for PEM Fuel Cells", 2018 IEEE Energy Conversion Congress and Exposition (ECCE), *2018 IEEE Energy Conversion Congress and Exposition (ECCE)*, pp. 3616-3622. IEEE, 2018.

P. Aboutalebi, A. Abbaspour, P. Forouzannezhad, A. Sargolzaei "A Novel Sensor Fault Detection in an Unmanned Quadrotor Based on Adaptive Neural Observer", *Journal of intelligent & robotic systems*, 90, no. 3-4 (2018): 473-484.