10-31-2022

A Link Between Brevetoxin Exposure, Oxidative Stress, and the Unfolded Protein Response in Human Lymphoblast Cells

Jordan Jobson
jjobs001@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

NEUROIMAGING FEATURE FUSION AND MULTIMODAL CLASSIFICATION OF ALZHEIMER'S DISEASE

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL AND COMPUTER ENGINEERING

by

Mehdi Shojaie

2022

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

This dissertation, written by Mehdi Shojaie, and entitled Neuroimaging Feature Fusion and Multimodal Classification of Alzheimer's Disease, 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.

_____
Mercedes Cabrerizo

_____
Armando Barreto

_____
Naphtali Rishe

_____
David A. Loewenstein

_____
Malek Adjouadi, Major Professor

Date of Defense: November 2, 2022

The dissertation of Mehdi Shojaie 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, 2022

DEDICATION

This dissertation is dedicated to my parents, for their endless love, support, and encouragement.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my major advisor Dr. Malek Adjouadi for his invaluable guidance, support, research ideas, and providing me with financial support throughout my doctoral studies and research. I am also grateful to Dr. Adjouadi for providing opportunities to collaborate with medical doctors and experts from the 1Florida Alzheimer's Disease Research Center (ADRC) and the Mount Sinai Medical Center. Specifically, I am greatly thankful to Dr. David Loewenstein and Dr. Ranjan Duara for their support on clinical matters and for their constructive comments and suggestions.

Furthermore, I would like to thank my other dissertation committee members Dr. Armando Barreto, Dr. Mercedes Cabrerizo, and Dr. Naphtali Rishe for their help, support, and valuable inputs.

Finally, I would like to acknowledge my colleagues at the Center for Advanced Technology and Education (CATE) for their wonderful collaboration and technical support and assistance.

This research is made possible thanks to the support by the National Science Foundation (NSF) under grants CNS-1920182, CNS-1551221, and CNS-2018611. The support of the 1Florida Alzheimer's Disease Research Center (ADRC) (NIA 1P50AG047266-01A1) is also greatly appreciated.

ABSTRACT OF THE DISSERTATION

NEUROIMAGING FEATURE FUSION AND MULTIMODAL CLASSIFICATION OF ALZHEIMER'S DISEASE

by

Mehdi Shojaie

Florida International University, 2022

Miami, Florida

Professor Malek Adjouadi, Major Professor

Alzheimer's Disease (AD) is a neurodegenerative disorder and the most prevalent cause of dementia. Early detection of AD is critical for enabling early intervention and for slowing its progression. This dissertation aims to present effective machine learning frameworks using multimodal biomarkers for the diagnosis of AD, specifically in the earliest stage. Moreover, a transfer-learning framework is proposed to transfer model learning knowledge from a source domain with a large amount of labeled data to a target domain with insufficient data for creating an ML model from scratch.

Accordingly, a feature ranking metric is formulated based on the mutual information index to assess the relevance and redundancy of regional biomarkers and improve the AD classification accuracy. The multivariate mutual information metric was utilized to capture the redundancy and complementarity of the predictors and develop a feature ranking approach. This was followed by evaluating the capability of single-modal and multimodal biomarkers in predicting the cognitive stage. Moreover, the Amyloid-Tau-Neurodegeneration (AT(N)) biomarker framework was used to explore the misclassified cases. The F1-score show that although amyloid-β deposition is an earlier event in the

disease trajectory, tau PET with feature selection yielded a higher early-stage classification score (65.4%) compared to the modalities of amyloid-β PET (63.3%) and MRI (63.2%). The support vector classifier (SVC) multimodal scenario with feature selection improved the F1-score to 70.0% and 71.8% for the early and late-stage, respectively.

As another research endeavor, an instance-based transfer-learning framework is presented based on the gradient boosting machine (GBM) to transfer knowledge from a source to a target domain. In our transfer learning version of GBM (TrGB) a weighting mechanism based on the residuals of the base learners is defined for the source instances. Consequently, instances with different distribution than the target data will have a lower impact on the target learner. Using the Mount-Sinai dataset as target domain, and the ADNI dataset as source domain, TrGB improved the classification scores by 4.5% for MCI diagnosis compared to the baseline. Also, for the early MCI vs. late MCI, using knowledge transfer from the NC vs. AD of the source domain, scores improved by 5%.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

SYMBOLS AND ABBREVIATIONS

Aβ	Amyloid-β
AD	Alzheimer's Disease
ADAS	The Alzheimer's Disease Assessment Scale
ADNI	Alzheimer's Disease Neuroimaging Initiative
ADRC	1Florida Alzheimer's Disease Research Center
APOE	Apolipoprotein E
AT(N)	Amyloid-Tau-Neurodegeneration
AUC	Area Under the Curve
CDR	Clinical Dementia Rating
CN	Cognitively Normal
CNN	Convolutional neural network
CSF	Cerebrospinal Fluid
DAM	Domain Adaptation Machine
EMCI	Early Mild Cognitive Impairment
ETC	ExtraTreesClassifier
FDG	Fluorodeoxyglucose
FS	Feature score
GBM	Gradient boosting machine
GM	Grey Matter
GRU	Gated recurrent units
LDA	linear discriminant analysis

LMCI	Late Mild Cognitive Impairment
LPP	Locality preserving projection
LSTM	Long-Short Term Memory
MCI	Mild Cognitive Impairment
METL	Multi-source ensemble transfer learning
MI	Mutual Information
ML	Machine Learning
MMI	Multivariate Mutual Information
MMSE	Mini-Mental State Exam
MRI	Magnetic Resonance Imaging
MSTL	Multi-source transfer learning
NIA	National Institute on Aging
PET	Positron Emission Tomography
PIB	Pittsburgh compound B
RGF	Radial basis function
RF	Random forest
RNN	Recurrent Neural Network
ROC	Receiver operating characteristic
ROI	Region Of Interest
SKB	SelectKBest
SPM	Statistical Parametric Mapping
SUVR	Standardized Uptake Value Ratio

SVM	Support Vector Machine
TrAdaBoost	Transfer learning AdaBoost
TrGB	Transfer learning gradient boosting
VOI	Volumes of interest
WM	White Matter
XGB	eXtreme Gradient Boosting

1. INTRODUCTION

1.1 Background

Alzheimer's Disease is a neurodegenerative disorder that impairs memory, thinking, and behavior. By 2050, the number of individuals with dementia is estimated to exceed 13.8 million in the U.S. according to Alzheimer's Disease Facts and Figures [1]. A misfolding and abnormal deposition of specific proteins in the brain is recognized as the pathological cause for the initiation and progression of this neurodegenerative disease. AD is irreversible, causing significant memory and behavioral issues. Therefore, researchers are keen to identify its earliest manifestations, even at the pre-symptomatic stage, to plan for and more effectively take advantage of emerging early treatment and therapeutic interventions. Thus, effective diagnosis of AD and its early stage, i.e., mild cognitive impairment (MCI), specifically using computer-aided methods, has attracted extensive attention in recent years [2]-[14].

Several biomarkers associated with the pathology of AD have been identified and studied by researchers for decades. Magnetic Resonance Imaging (MRI) as a structural indicator for brain atrophy, measures of tau and amyloid-β (Aβ) from Cerebrospinal Fluid (CSF), and Aβ accumulation from regional Positron Emission Tomography (PET) and hypometabolism from fluorodeoxyglucose (FDG) PET are among the most remarkable biomarkers for AD. In recent years, several tau PET tracers such as (11)C-PBB3, (18)F-AV1451, and (18)F-THK have been developed, which enable in-vivo visualization of tau pathology in brain regions. Tau imaging can help to facilitate disease staging and diagnosis. Compared to Aβ, tau is a delayed event and is more related to cognitive decline [15], [16]. The interrelatedness of these two biomarkers has been extensively studied [17]-[21].

Moreover, the temporal ordering of biomarkers provides added insight into AD staging. Based on such biomarkers ordering, a disease progression score has been defined in [22]. Biomarkers of Aβ plaque, i.e., amyloid PET and CSF Aβ, represent the initiating events of AD that happen during the cognitively normal stage. On the other hand, biomarkers of neurodegeneration, including MRI, FDG-PET, and CSF total tau, are later events that correlate with cognitive decline [23]. Besides the pathological biomarkers, there are other contributing variables in AD diagnosis, such as risk factors (age, gender, and apolipoprotein E gene with the e4 allele (APOE e4)), and protective factors (cognitive reserve, brain resilience, and resistance). The variability of these factors, namely age, gender, APOE e4 genotype, and year of education between AD subtypes, can be used and investigated to address the disease heterogeneity to some extent.

In various fields, such as computer vision, natural language processing, speech recognition, and intelligent humanoids, machine learning has achieved promising advances. In the medical domain, machine learning is used to create learning models as complementary tools for disease diagnosis [24], [25]. However, collecting and labeling medical data is costly and labor-intensive. Thus, some medical datasets may not have enough data required for training a robust machine learning model. On the other hand, combining various datasets may introduce new challenges such as differences in the marginal and conditional distribution of the data due to disparity in imaging machines and tools, data collection policies and standardization, and characteristics of the participants. In machine learning, it is often assumed that the feature and label spaces and their distribution are the same for the training and testing datasets. If the data distribution is not the same, a trained learner on a source dataset may perform poorly on a target dataset. On

the other hand, the labeled data of the target domain may not be large enough to train a separate model. To handle this conflicting challenge, transfer learning is used to effectively transfer the information between domains while taking into consideration potential discrepancy in data distribution. This will make it possible to build predictive models in a target dataset with a limited amount of data [26]. In general, transfer learning can be used in a scenario where there is a link between the two learning tasks. If the connection between the domains is misleading or misinterpreted by the model, negative transfer may occur where information transfer has a negative impact on the target learner [27].

1.2 Research Purpose

The objective of this research work is to develop computer-aided AD diagnosis approaches using multimodal neuroimaging data including MRI, Aβ PET, and tau PET. Tau imaging is a relatively new technology which has not yet been studied thoroughly in the machine learning-based AD diagnosis domain in comparison to other modalities. Various biomarkers can thus be studied either individually to gauge their unique merit but also in a multimodal manner to discover how they complement one another for enhanced classification and disease prediction. The spatial variation of tau and Aβ proteins in different stages of the disease is investigated. A feature selection methodology will be proposed to determine the relevancy and redundancy of the features. Then, machine learning architectures are utilized for multiclass and binary classification of MCI and AD patients. The structure of the model is designed so that the inter-modal information can be extracted effectively. To interpret the diagnosis results, the relationship between biomarkers profiles and the cognitive stage is also examined.

The machine learning-based diagnosis framework was also expanded to a secondary target dataset. However, since the amount of data in this dataset is not sufficient to build a separate classification model, a transfer learning approach is developed to transfer information through instances from a larger source dataset to the smaller target dataset. The transferred instances will facilitate the training process of the target learner and improve its performance.

1.3 Research Problem

In the procedure of developing machine learning models for AD classification and feature selection for determining disease-prone areas, the following steps are undertaken:

- *Biomarkers data exploratory*: This analysis is to identify the most effective biomarkers and to investigate the fusion of multimodal data to enhance model performance. This step is essential for studying the temporal ordering of biomarkers and detecting the early biomarkers. Extracting various biomarkers profiles using defined cut-points and explaining them in the context of cognitive status can also be helpful.

- *Relevancy and redundancy of regional features*: Identifying the relevance of the regional features for each modality in different stages of the disease could enhance the prospects for classification and prediction purposes. Also, in the process of integrating features, their redundancy as well as their complementarity should be considered to optimize the machine learning design and hence the classification/prediction process, while minimizing the computational requirements.

- *Designing the machine learning structure*: In seeking that optimal design structure, there is need to identify the appropriate machine learning models for the given

classification/prediction problem. Developing such a structure to integrate multimodal regional data can effectively model both inter-modal and intra-modal relationships. Designing machine learning architectures to enable transfer learning to combine datasets with different data distributions is key in developing machine learning models that can be applied across data centers and even across disciplines.

1.4 Theoretical Perspective and Literature Review

While each neuroimaging modality provides distinct features and measures for AD diagnosis, their fusion consolidates their unique strengths when using effective machine learning and deep learning models [28]-[32]. In retrospect, few multimodal studies include tau imaging for computer-aided diagnosis of AD.

An initial step required for the machine learning-based diagnosis is the optimal data representation through a feature extraction procedure. Feature extraction methods can be categorized as voxel-based, region of interest (ROI)-based, and patch-based techniques. Among them, ROI-based features are more common due to their consistency and lower dimensionality [28], [33]. In AD studies, the sample size is typically small, and the dimensionality of voxel-based and even ROI-based features is high. This makes it difficult for the machine learning model to generalize to unseen data while avoiding overfitting. Therefore, feature selection methods are used to remove the irrelevant features and reduce the dimensionality of the data.

In some feature selection methods, the selection process is embedded in the learning algorithm, and the model accuracy or loss is then used to evaluate different subsets of features. With the use of these methods, an optimized combination of features can be

achieved; however, these approaches are subject to the *curse of dimensionality*. Another category of techniques known as filter methods uses a criterion such as Pearson's correlation, ANOVA, t-test, chi-square test, and mutual information, among others, to evaluate the many features and determine their relevance to the target variable [34], [35]. In [36], the similarity between samples was computed, and their consistency metrics have been used for multimodal feature selection. In [37], a feature selection method was developed based on the Receiver Operating Characteristic (ROC) curve for each Volumes-Of-Interest (VOI), where the classification true positive rate is plotted vs. the false positive rate using only that specific VOI. In [38], the linear discriminant analysis (LDA) and locality preserving projection (LPP) learning methods have been combined with a sparse regression model to determine discriminative features. Most filter methods use univariate metrics in which features are evaluated independently, and the interaction between them is often overlooked. Also, filter methods focus mainly on the linear relationship between variables, and any nonlinear dependencies are often neglected.

In terms of associations between variables, there exist some research endeavors for incorporating the correlation and redundancy of the features. However, due to the nature of the used metrics, these approaches are mainly unsupervised, and the detected relationships are not necessarily connected to the target variable and hence may not be valuable concerning the classification problem. Another group of methods uses embedded regularization for sparse feature learning in which the interaction of all variables is taken into consideration [39]-[41]. However, in these models, the variable selection is less interpretable, limiting the flexibility and ability to further explore the discriminative features.

To present a biological definition of AD, biomarkers are pathologically grouped into three classes. This scheme is known as AT(N) with "A", "T", and "(N)" representing Aβ, tau, and neurodegeneration biomarker groups, respectively. Based on this system, each biomarker class is labeled as positive or negative through defined cut-points to determine the overall pathology status [42]. The so-called AT(N) framework attempts to reflect the interactions between neuropathological changes (characterized by biomarkers profiles) and the cognitive stage (determined clinically through symptoms). This framework can be extremely useful when interpreting results of a computer-aided diagnosis model.

1.5 Transfer Learning – A Retrospective

Transfer learning approaches can be classified into instance-based, feature-based, or model-based techniques. Instance-based approaches focus on transferring knowledge through the source domain instances while using a weighting scheme as a filtering criterion. In feature-based transfer learning, the source features are transformed, and a new feature space is built. The aim of the model-based transfer learning is to take a pre-trained model built on a primary dataset and fine-tune and use it for a smaller dataset [43]. **Tr**ansfer **Ada**ptive **Boost** (TrAdaBoost), which is an extension of the AdaBoost algorithm, has been presented in [44] as an instance-based method. In this iterative approach, the training data of the source and target domains are combined, and a model is trained. In each iteration, if an instance from the target domain is misclassified, a higher weight is assigned to it, while the misclassified instances from the source domain will receive a lower weight for the next step. Thus, the different-distribution training instances from the source domain will have a lower impact on the final model. An extended version of TrAdaBoost for multi-source transfer learning scenarios has been developed in [45] and is called MsTrAdaBoost.

Another multi-source transfer learning (MSTL) approach has been proposed in [46] to facilitate the utilization of knowledge from multiple source domains. In [47], a multi-source ensemble transfer learning (METL) approach is presented, which consists of a single-source tri-transfer learning and a multi-source ensemble learning.

Furthermore, deep learning models such as convolutional neural networks (CNNs) have been extensively used for the automatic extraction of discriminative features and classification. Since deep learning requires a massive amount of data points, they are ideal candidates for transfer learning [48]. For such situations, a part of the network will be transferred in a model-based manner. In a group of studies, the convolutional layers of feature extraction that are pre-trained on a general dataset are transferred for a specific task of disease diagnosis. In contrast, the classifier and fully connected layers are designed and trained for the desired classification task [49], [50]-[53]. Another approach to realizing model-based transfer learning is to add a pre-trained model from a source domain to the objective function of a target learner and transfer the source knowledge during the target training procedure. A version of this approach has been proposed in a Domain Adaptation Machine (DAM) in [54].

Progression of AD can be effectively investigated through longitudinal studies. Recurrent Neural Networks (RNNs) are perfect candidates for such studies due to the intrinsic power in learning long-short term dependencies of sequenced data. These RNNs can share information between series of data points through an additional hidden set of parameters. RNNs have been implemented in modeling the progression patterns of chronic diseases [55], [56]. In [57], Nguyen et al. trained an RNN-LSTM network over a seven-year period to predict multiple AD biomarkers for one subsequent time point. In another

study, Wang et al. applied an RNN architecture with LSTM cells to predict the global staging of the Clinical Dementia Rating (CDR) score of the next visit using previous records [58]. Aghili et al. utilized LSTM and GRU models to classify AD subjects using longitudinal records of data over an 11-year period [59].

2. FEATURE ENGINEERING

2.1 Introduction

To reduce the model complexity and enhance its performance, removing redundant and extraneous features by selecting the most informative ones remains a critical step [36], [60]-[62] in the development and design of any machine learning model. Also, feature selection can be used to understand the process under study by identifying those disease-prone regions that contribute best to AD diagnosis and disease progression.

In this study, the objective is to come up with a subset of features with minimum size and maximum possible information about the class variable. This can be achieved by preserving the most relevant features and dismissing the irrelevant and the redundant ones. Redundant features may not necessarily damage the system's performance. However, to limit the feature space size and complexity, it is beneficial to remove the redundant features and keep the complementary ones to maximize the total amount of relevant information.

Within this context, a new approach is thus proposed based on the Multivariate Mutual Information (MMI) criterion. We attempted to handle feature redundancy and complementarity in a supervised manner where the shared information between features is evaluated in terms of its capability to predict the target variable. A feature score is thus defined based on the features' relevance and redundancy, and the features are subsequently ranked based on their scores.

2.2 Materials and method

2.2.1 Participants

The clinical data used for our analysis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). ADNI was launched in 2003 as a public-private partnership, directed by Principal Investigator Michael W. Weiner, MD. The primary objective of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease. For up-to-date information, see www.adni-info.org.

In this study, the data were collected from three modalities in the ADNI 3 cohort, including amyloid PET (agent: (18)F-AV45), tau PET (agent: (18)F-AV1451), and MRI. For each participant, all modalities have been collected from the same visit. The MRI scan is a T1 weighted image that has gone through preprocessing steps, including gradient wrapping, scaling, B1 correction, and inhomogeneity correction. For the florbetapir and flortaucipir data, four preprocessing steps have been followed, including co-registered dynamic, averaged, standardized image and voxel size, and uniform resolution. T1 MRI scans have been processed through FreeSurfer for skull-stripping and segmentation of cortical and subcortical regions. In the next step, florbetapir and flortaucipir images have been co-registered to the subject's MRI from the same visit. Finally, volume-weighted florbetapir and flortaucipir average are defined in each cortical and subcortical region of interest, and regional Standardized Uptake Value Ratio (SUVR) is then calculated. More information about the preprocessing steps and processing methods can be found at ida.loni.usc.edu. The florbetapir ((18)F-AV45) dataset analysis comprises reference region

options of the whole cerebellum, cerebellar grey matter, and brain stem in addition to cortical and summary of SUVR measurements. The participant demographics and Mini-Mental State Examination (MMSE) score for each group (mean and standard deviation) are reported in Table 2.1.

Table 2.1 Participant demographics and mini-mental state examination (MMSE) score for different diagnosis groups of the ADNI3 cohort. P-value is reported between MCI-CN and AD-CN populations.

Groups	Subject (f/m)	Age (year) [P-value]	Education (year) [P-value]	MMSE [P-value]
CN	277(153/124)	71.80±5.70 [-]	16.67±2.47 [-]	28.63±2.12 [-]
MCI	378(155/223)	71.26±7.66 [0.179]	16.25±2.61 [0.027]	26.87±4.20 [<0.001]
AD	67(26/41)	73.41±8.78 [0.075]	16.43±2.35 [0.290]	22.37±2.39 [<0.001]

Figure 2.1 illustrates the distribution of average SUVRs (among all regions) for the sample set. Since not all participants have undergone all tests, the dataset contains multiple instances with missing values which are dropped in some scenarios depending on the objective of the analysis.

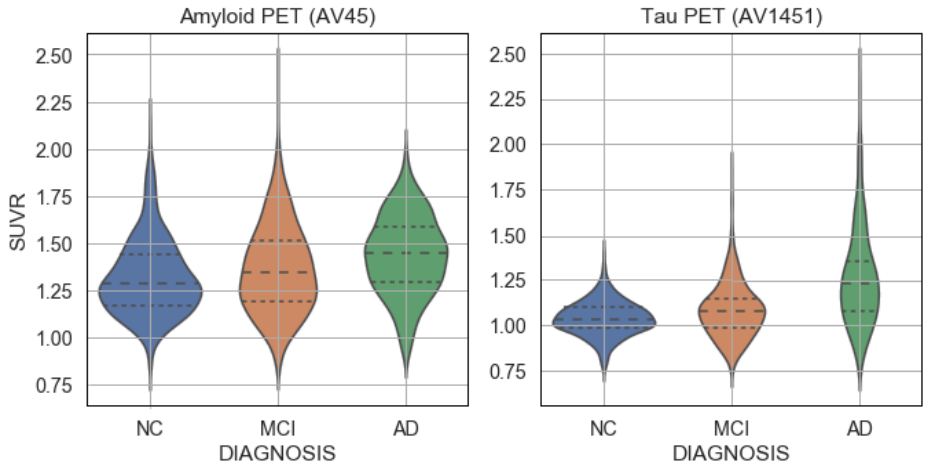


Figure 2.1 Distribution of the mean value of amyloid-β and tau SUVRs in each disease group for ADNI3 cohort participants; CN: Cognitively Normal, MCI: Mild Cognitive Impairment, AD: Alzheimer's Disease

In this study, different types of variables, including cortical thickness and SUVR values, non-tissue SUVR values, and AD risk factors, were used as features for the machine learning algorithm. In the preprocessing stage, the feature set is normalized to a common scale before feeding it to the classification model. It is worth noting that the SUVR values in non-brain areas represent off-target binding by the ligand and are not related to AD pathophysiology. Such SUVR values could still be potentially beneficial for the machine learning-based classification task despite the fact that they are not interpretable as biomarkers of AD.

2.2.2 Method: Feature selection

The high dimensionality of multimodal regional AD data relative to the sample size can diminish the model performance. The purpose of feature selection is to find a feature subset that yields an optimal classification score. This selection process can also help to enhance the generalization ability and interpretability of the model.

To determine the relevance of a feature, univariate filter-based feature selection measures can be used. With such measures, the relationship between each feature and the target variable is evaluated individually. One of the most common criteria for this task is the Pearson correlation coefficient which is a number between [-1,1], with +1, -1, and 0 representing maximum linear correlation, maximum inverse linear correlation, and no linear correlation between the two variables, respectively. Other univariate criteria include mutual information, ANOVA test, and Chi-squared test, whose performance may vary depending on the type of the input and output variables (continuous or categorical variable). Mutual Information (MI) is a powerful statistical metric that measures common information between random variables and is relatively robust to the data type. Unlike the correlation measure, MI can also detect nonlinear relationships between variables. Moreover, it can be extended to more than two variables to determine the redundancy of multiple variables [62]. In this study, a methodology is proposed to rank features based on pairwise redundancy and complementarity of features using Multivariate Mutual Information (MMI).

MI between two discrete random variables is defined as:

$$I(x;y) = \sum_x \sum_y p(x,y).log\frac{p(x,y)}{p(x)p(y)} \quad (2\text{–}1)$$

where x and y are random variables and p(.) is the probability of a random variable. MI is zero when x and y are independent and is positive when there is common information between them.

At first, MI was calculated between each feature and its target variable. This determines the relevance of each feature. Next, to incorporate the interaction of features,

MI was calculated between a subset of features and a target variable as I(S; y), where S is a subset of features and y is the target. For the case of a subset of two features (S={x_1, x_2}), the relationship between MI of S and y (I(x_1, x_2; y)) and MI of each feature and y (I(x_1; y), I(x_2; y)) is defined as follows:

$$I(x_1,x_2;y) = I(x_1;y) + I(x_2;y) - I(x_1;x_2;y) \quad (2\text{–}2)$$

where the three terms on the right side can be calculated using (1). Based on (2), the amount of information that (x_1,x_2) have about y can be defined as the sum of the common information of x_1 and y (I(x_1; y)) plus that of x_2 and y (I(x_2; y)) minus the intersection of the first two terms, which is the common information of all three variables x_1, x_2 and y (I(x_1; x_2; y)). The last term is known as the Multivariate Mutual Information (MMI), which determines the shared information between multiple variables and is defined as follows:

$$I(x_1;x_2;y) = \sum_{x_1}\sum_{x_2}\sum_{y} p(x_1,x_2,y).log\frac{p(x_1,x_2,y)}{p(x_1)p(x_2)p(y)} \quad (2\text{–}3)$$

When MMI is positive, there is redundancy between x_1 and x_2, and the information of a subset of them is less than the sum of their individual information. On the other hand, when MMI is negative, x_1 and x_2 carry complementary information about y, and the information of x_1 and x_2 combined is more than the sum of their individual information. Therefore, in (2), the interaction of features is considered through the MMI term, which can be treated as a measure of redundancy and complementarity.

To rank the features, a metric is defined for each feature based on the MI between that feature and the target variable and the redundancy or complementarity of that feature with every other feature. This new metric is defined as follows:

$$FS_i = I(x_i; y) - \alpha \sum_{j \neq i} {}_j \, I(x_i; x_j; y) \quad (2\text{–}4)$$

where FS_i is the score of the i^{th} feature, with α being a constant. The first term is the MI of the i^{th} feature and the target variable, and the second term represents the pairwise interaction (redundancy/complementarity) of the i^{th} feature and all other features, which can consist of positive and negative elements. When α is zero, the interaction term is ignored, and the feature scores only depend on the individual scores. As α increases, a larger weight is assigned to the redundancy term so that the overall score of redundant features decreases while that of complementary ones increases. To select the value of coefficient α, the classification experiment was conducted using different values of α, and the optimal value was determined as the one associated with the highest classification score. The FS score was then calculated for all features, and the top features were determined accordingly. To evaluate different scenarios, first, the top features were detected for each individual modality to find the prominent regions based on each biomarker. Then, the process was repeated for the multimodal data so that the top regions in terms of all modalities combined were identified. Also, the importance of specific regions and biomarkers at various stages of the disease was evaluated. In the next step, to prove the effectiveness of the new metric for feature selection, multiple classification scenarios were implemented.

2.3 Results and Discussion

Various feature selection approaches were implemented under multiple classification scenarios. At first, conventional univariate criteria and methods, including Correlation coefficient, ANOVA (SelectKBest), ExtraTreesClassifier, and univariate mutual

information, were implemented. For the amyloid and tau PET modalities and the three-class classification case (CN/MCI/AD), the heatmap of the feature scores based on the abovementioned metrics is shown in Fig. 2. A total number of 110 features (two features per region for left and right hemispheres) have been included in this analysis. As seen, entorhinal, inferior parietal, inferior temporal, amygdala, and bankssts are among the top features based on tau PET, while regions like frontal pole and accumbens are more prominent based on amyloid PET.

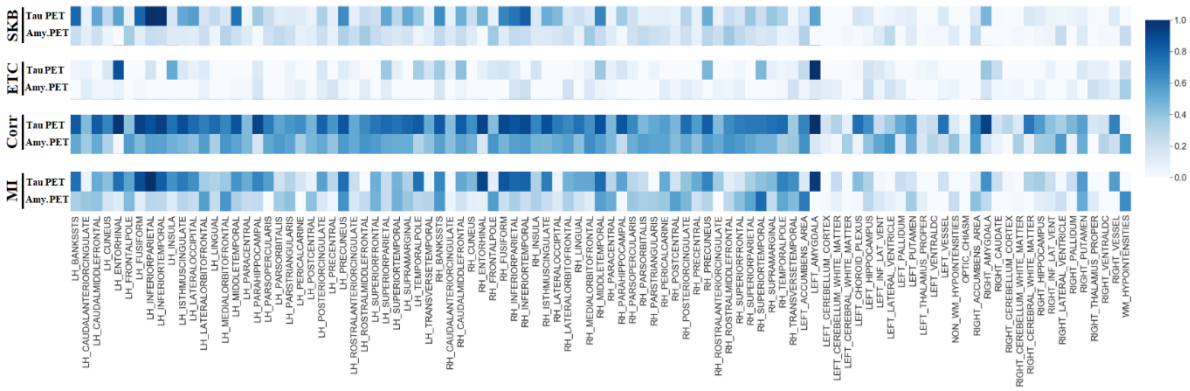


Figure 2.2 Regional feature importance scores for amyloid PET SUVRs (AV45) and tau PET SUVRs (AV1451). The feature scores were determined using four filter-based feature selection measures, namely, SelectKBest (SKB-ANOVA), ExtraTreesClassifier (ETC), correlation coefficient (Corr), and mutual information (MI), as shown in the vertical axis. For each region shown in the horizontal axis, one feature is defined for amyloid SUVR and one for tau SUVR. The value of feature scores is normalized between 0 and 1 and is illustrated by the color intensity of their corresponding box in the figure. Features with larger scores are more informative for the classification task. Based on the results, amyloid SUVRs including entorhinal, inferior parietal, inferior temporal, amygdala, and bankssts and tau SUVRs including frontal pole and accumbens are among the top features.

Next, the proposed MMI-based feature selection method was implemented. Using equation (3), pairwise MMI was calculated for all features, and the results are presented as a heatmap in Figure 2.3. Again, the CN/MCI/AD case based on the amyloid and tau PET modalities is considered here. In the heatmap, the diagonal elements show the amount of

information that each feature has about the target variable. The brighter the color of a square, the more relevant is that particular feature. The non-diagonal elements show the degree of redundancy or complementarity of feature pairs concerning the target variable. The darker the color, the higher the redundancy, and the lower the complementarity.

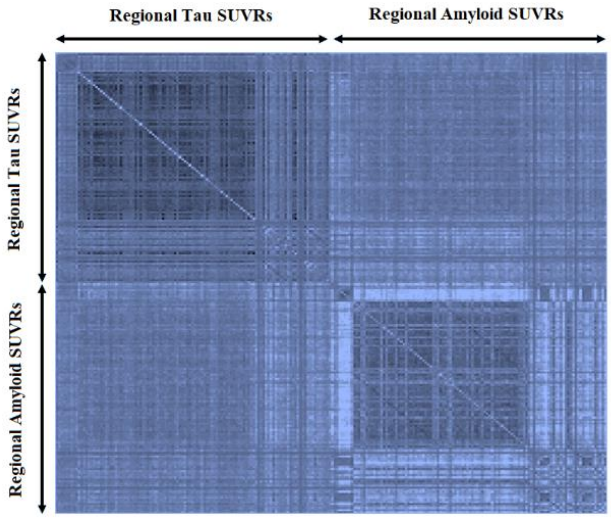


Figure 2.3 Heatmap of multivariate mutual information (MMI) between pairwise amyloid and tau SUVR values given the class variable (y), calculated using equation (3). The diagonal elements represent the amount of information that each individual feature carries about the target variable. Brighter colors correspond to a higher amount of information. For non-diagonal elements, a positive MMI value is an indication of redundant information between two features, which corresponds to darker colors in the heatmap. On the other hand, complementary features have a negative MMI represented by brighter colors in the heatmap. As seen, more pairwise redundancy (more dark non-diagonal elements) exists for inside-modality features compared to between-modality features.

To select the most relevant and informative features, both the individual scores (diagonal) and the mutual scores (non-diagonal) should be considered as described in the method section. The feature scores (FS) were calculated using equation (4). As indicated earlier, for each feature, the summation of the second term of the equation represents the interaction of that feature with every other feature. The summation terms are equivalent to each row or column of the heatmap of Figure 2.3. The heatmap of the top 30 features based

on the proposed FS-score is illustrated in Figure 2.4 for different values of α. For α=0, the score of a given feature solely depends on the feature's relevance. As seen in Figure 2.4, in this case, top features include highly relevant (brighter diagonal) but possibly redundant features (darker non-diagonal) at the same time. For higher values of α, the redundancy term comes into play so that more redundant features are removed from the list of the top features. This results in selecting features with brighter non-diagonal elements (less redundant), as shown in Figure 2.4 for higher values of α. This is a trade-off between feature relevance and redundancy, which is controlled by adjusting parameter α.

Table 2.2 Top features (amyloid-β and tau SUVRs) based on the proposed feature ranking method. The SUVR values were ranked using the calculated feature scores, and the top amyloid-β and tau SUVR features are presented. Top features are more informative for the AD diagnosis classification task.

Tau PET	Left entorhinal	Left vessel	Third ventricle
	Left amygdala	Left inferior temporal	Right entorhinal
	Left middle temporal	Right amygdala	Right inferior temporal
Amyloid-β PET	Left medial orbitofrontal	Left rostral anterior cingulate	Right medial orbitofrontal
	Left accumbens area	Left hippocampus	cingulate cortex anterior
	Left frontal pole	Right accumbens area	cingulate cortex mid anterior
	Left lateral ventricle	Right lateral ventricle	cingulate cortex posterior
	Left inferior lateral ventricle	Right frontal pole	

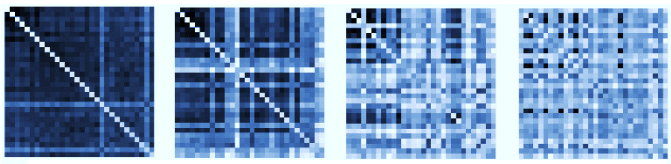


Figure 2.4 Heatmap of top 30 features based on the FS-scores for different values of parameter α. For α=0, the redundancy term is ignored, and the features are selected solely based on their relevance. In this case, dark non-diagonal elements of the heatmap represent more pairwise redundancy between features. For higher values of α, feature redundancy is decreased, and bright non-diagonal elements show less pairwise feature redundancy and more complementarity.

It is worthwhile to add that too large values of α should be avoided since, in this left inferior lateral ventricle situation, valuable features might be dropped only because they have some dependency on other features. For the specific case of α=0.005, the top features (amyloid-β and tau SUVRs) are listed in Table 2.2. Finally, the resulting scaled feature scores for the amyloid and tau SUVRs for different stages of the disease are represented in Figure 2.5.

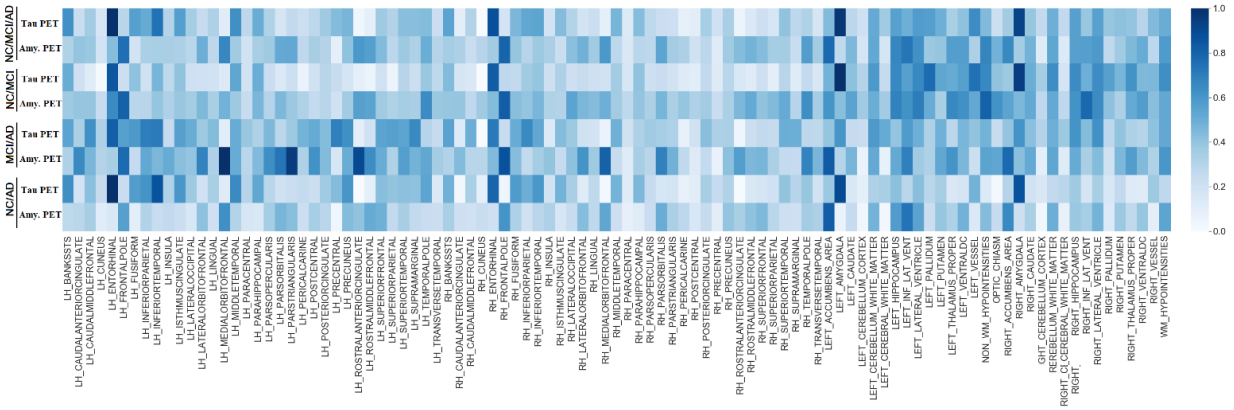


Figure 2.5 Regional feature importance scores for amyloid PET SUVR (AV45) and tau PET SUVR (AV1451) based on the proposed feature selection method. As a supervised approach, the features scoring procedure was performed for four different classification tasks, including CN/MCI/AD, CN/MCI, MCI/AD, and CN/MCI/AD as shown in the vertical axis. For each region shown in the horizontal axis, one feature is defined for amyloid SUVR and one for tau SUVR. The value of feature scores is normalized between 0 and 1 and is illustrated by the color intensity of their corresponding box in the figure. Features with larger scores are more informative for the classification task. For tau SUVRs, entorhinal and amygdala were among the top features for all classification tasks, while pallidum and hippocampus were more informative for the CN/MCI case, and inferior parietal, inferior temporal, precuneus, and precentral for the MCI/AD case. On the other hand, for amyloid SUVRs, top features include frontal pole for all classification tasks, in ferior lateral ventricle for the CN/MCI, and medial orbitofrontal, pars triangularis, and rostral anterior cingulate for the MCI/AD.

2.4 Conclusion and Discussion

The objective of this chapter was to determine the dependencies between different features. For the model variables, including amyloid and tau PET SUVR values and cortical thickness, a trade-off was made between variables relevance and redundancy using an information theory-based metric. The advantage of the proposed approach is to incorporate the effect of feature complementarity and redundancy to maximize the total amount of information in the feature set. By incorporating a moderate redundancy coefficient into the calculations for tau SUVRs, entorhinal and amygdala were among the top regions for all stages of AD, with amygdala being most informative for the CN/MCI case. Abnormal tau deposition in these regions is known as a biomarker for preclinical AD by previous studies [18], [20], [63]. It is reported in the literature that amygdala shows early atrophy independent of amyloid deposition, and it might be related to neurofibrillary tangles instead [64], [65]. Other prominent regions include pallidum and hippocampus based on tau PET for CN/MCI case, and inferior parietal, inferior temporal, precuneus, and precentral for the MCI/AD case. It is stated in [66]-[68] that tau burden in these specific ROIs is correlated with cognitive decline. On the other hand, for amyloid PET SUVRs, frontal pole for all stages, and inferior lateral ventricle for the CN/MCI case, and medial orbitofrontal, pars triangularis, and rostral anterior cingulate for the MCI/AD case are among the more prominent variables. These findings are consistent with previous studies [69]-[71].

By incorporating the effect of redundancy and synergy, some features experienced a score change. For instance, the score of frontal pole amyloid SUVR (but not tau SUVR) for the early stage increased significantly, so this region is considered a complementary variable for the classification task. This is in agreement with the literature [64], [72], where

it is reported that the frontal pole shows early amyloid deposition while atrophy and tau deposition are later events. Some amyloid and tau SUVR values that experienced a boost in their score include the hippocampus, inferior lateral ventricle, and lateral ventricle, which are known to be critical for AD diagnosis in previous studies. On the other hand, a score drop was seen in some of the tau SUVRs, including fusiform, inferior parietal, inferior temporal, isthmus cingulate, orbitofrontal, middle temporal, precuneus, and bankssts (these are cortical areas around superior temporal sulcus). A lower score does not necessarily disqualify a feature. Instead, the model tries to replace the most redundant features with a possibly less relevant but complementary one so that additional information is added to the analysis.

3. CLASSIFICATION AND INTERPRETATION USING THE ATN FRAMEWORK

3.1 Introduction

In recent years, artificial intelligence has proved to be a promising tool for diagnosing and predicting the trajectory of the disease. In this chapter, neuroimaging data, including MRI, Amyloid-β PET, and tau PET data from the ADNI cohort were used in this multimodal study. The effect of modalities on the disease staging was evaluated both individually and combined. Machine learning models, including Support Vector Machine (SVM), Random Forest (RF), and eXtreme Gradient Boosting (XGB), were used for the classification of different stages of the disease and the effect of the proposed feature selection method on the classification performance was evaluated. Before implementing the classification task, the feature space was scaled in the range between zero and one. To evaluate the models and also to optimize the models' parameters, k-fold cross-validation was used. In order to prevent data leakage between these two tasks, a nested cross-validation technique was implemented.

Lastly, the AT(N) biomarkers framework was used to investigate the interconnection between the biomarkers' profile and the cognitive stage to assess the classification performance degradation due to biomarker insufficiency.

3.2 Method: Classification

In this section, machine learning models were used for AD diagnosis at different stages using single-modality and multimodality data. The scaling estimator was built solely based on the training data (to avoid data leakage from the test set) and was applied to each feature individually in both training and test sets so that each feature is in the [0-1] interval.

The models used for the classification task include Support Vector Classifier (SVC), Random Forest (RF) of decision trees, and eXtreme Gradient Boosting (XGB). SVC is a classifier that attempts to categorize data points based on their classes in a high-dimensional space by a hyperplane. By mapping the data points onto a higher-dimensional space, SVC can classify non-linearly separable data using nonlinear kernels like polynomial and radial basis function (RBF). To alter the bias and variance of the model, the regularization parameters C and gamma of the SVC can be adjusted. The parameters control the trade-off between the training accuracy and model generalization ability for the testing stage. As the next model, the RF algorithm relies on the key concept of decision trees and leverages the ensembling and voting mechanisms to enhance the classification and prediction accuracy while preventing overfitting. The model parameters include the number of trees, sample size, maximum depth of each tree, and the maximum number of features used for each split. XGB, on the other hand, is a learning technique that consists of an ensemble of weak learners, such as decision trees, that operate in a sequence where each subsequent learner attempts to correct the errors of the previous learner. The number of trees, the maximum depth of a tree, and the sample size for each step are among the XGB control parameters. Nested cross-validation was performed for hyperparameter optimization, and an outer cross-validation was used for validation and reporting the model scores. The structure of the data for the classification task is shown in Figure 3.1. Multiple single modality and multimodality experiments were performed for binary and multiclass classification. A similar set of experiments were then implemented after applying the proposed feature selection approach. Finally, to include the risk and protective factors in

the analysis, covariates including age, APOE e4, gender, and education level were integrated into the feature set, and the classification process was repeated.

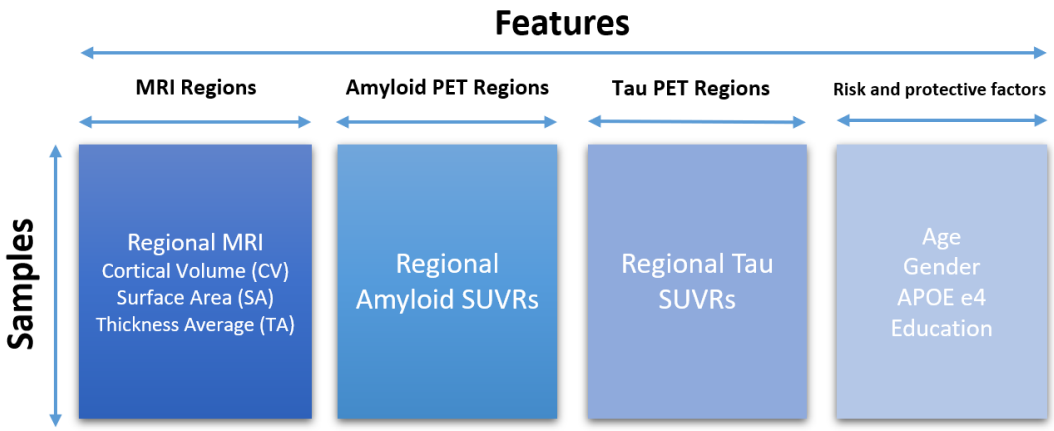


Figure 3.1 Structure of the used data for the classification process.

3.3 Interconnection between AD neuropathology and cognitive stage

In this chapter, MRI and PET scans have been used for automatic classification and prediction of the cognitive stage. However, the classification task remains challenging due to the heterogeneity of the disease. A critical factor that can degrade the model performance is the lack of sufficient biomarkers that are informative enough to perfectly determine the cognitive stage. We tried to explore the available biomarkers to investigate the performance limitation imposed by the dataset.

Due to biomarker insufficiency, cognitive symptoms are not perfectly linked to AD neuropathological changes measured by available biomarkers. Simply put, symptoms are not specific to AD, nor do abnormal AD biomarkers guarantee the existence of symptoms. Neuropathologic changes in Alzheimer's disease are determined by postmortem inspections and measured in vivo through biomarkers. Clinical AD, on the other hand, is defined based on the cognitive stage and is measured through the symptoms' manifestation.

A percentage of individuals with clinical AD do not have postmortem evidence of AD pathology.

Similarly, some individuals in the cognitively normal elderly group show signs of AD pathology at autopsy. This may result in false-negative and false-positive outcomes in our classification task. To study this effect, we investigated the available biomarkers and their corresponding cognitive stage based on the AT(N) biomarker profile system introduced in [42]. The AT(N) framework of the National Institute on Aging—Alzheimer's Association is an effort toward investigating the interaction between AD neuropathology and cognitive status. In this biomarker grouping system, the biomarkers are classified into three categories based on their underlying pathologic process. The label "A" represents amyloid PET and CSF Aβ as biomarkers of cortical Aβ, "T" denotes tau PET and CSF phosphorylated tau (P-tau) as biomarkers of fibrillar tau, and neurodegeneration is labeled as "(N)" measured by CSF total tau (T-tau), FDG PET, and MRI.

The imaging and CSF biomarkers are expressed in continuous values; however, in certain situations such as research studies and treatment trials, a binary grouping of biomarkers (positive/negative) may be preferred. To achieve such types of positive/negative results, appropriate cut-points are defined for each biomarker. For florbetapir (AV45) SUVR cut-points, we adopted the values reported in [73]. Summary SUVR is defined as the weighted average of florbetapir uptake in lateral temporal and parietal, lateral and medial frontal, anterior, and posterior cingulate normalized by the uptake in the whole cerebellum. Then, a cut-point of 1.11 is applied to this summary SUVR, which is equivalent to the 95th percentile of the biomarker distribution of the young control normal group. For tau PET SUVRs and MRI cortical thickness, the cut-points

determined in [74] by Clifford R. Jack Jr. were used. A tau PET summary SUVR is defined based on the volume-weighted average of the SUVR in inferior temporal, middle temporal, entorhinal, amygdala, parahippocampal, and fusiform ROIs normalized to the cerebellar crus grey. For the tau PET summary SUVR, cut-points of 1.19 and 1.32 are defined based on the specificity method (the 95th percentile of the biomarker distribution of the young control normal individuals) and the accuracy of impaired versus age-matched control normal method, respectively. From MRI, the surface-area weighted average is determined for the cortical thickness in entorhinal, inferior temporal, middle temporal, and fusiform regions. Cortical thickness cut-points of 2.69 and 2.57 mm are selected, respectively based on specificity and accuracy methods which were also used in the tau PET case.

Based on the defined cut-points, various biomarker profiles can be identified in the AT(N) framework. These biomarkers grouping and their relationship with the cognitive stages are shown in Table 3.1. As seen in the table, the A-T-N- group represents individuals with normal AD biomarkers. Participants with amyloid positive but normal tau pathology and neurodegeneration biomarkers (A+T-N-) are tagged as "Alzheimer's pathologic change." Those with evidence of amyloid deposition along with tau pathology and regardless of neurodegeneration condition (A+T+N+/-) are considered to belong to the "preclinical Alzheimer's disease" group. Amyloid-negative individuals with abnormal tau or neurodegeneration biomarkers (A-T-N+, A-T+N-, A-T+N+) are defined as "suspected non-Alzheimer's pathology change". Finally, the A+T-N+ category represents simultaneous "Alzheimer's pathologic change" and "non-AD neurodegeneration". Although the biomarker signature carries some information about the cognition status, each biomarker profile can belong to any cognitive stage.

Table 3.1 Interaction between clinically diagnosed cognitive stage and AT(N) biomarkers [42]

		Cognitive stage (Clinical diagnosis)		
		Cognitively normal (CN)	**Mild Cognitive Impairment (MCI)**	**Dementia**
Biomarker Profile	**A-T-N-**	Normal AD biomarkers, and CN	Normal AD biomarkers with MCI	Normal AD biomarkers with dementia
	A+T-N-	AD pathologic change, and CN	AD pathologic change with MCI	AD pathologic change with dementia
	A+T+N- **A+T+N+**	Preclinical AD with no cognitive impairment	AD biomarkers with MCI	AD biomarkers with dementia
	A+T-N+	Alzheimer's and concomitant suspected non-Alzheimer's pathologic change, and CN	Alzheimer's and concomitant suspected non-Alzheimer's pathologic change with MCI	Alzheimer's and concomitant suspected non-Alzheimer's pathologic change with dementia
	A-T+N- **A-T-N+** **A-T+N+**	non-Alzheimer's pathologic change, and CN	non-Alzheimer's pathologic change with MCI	non-Alzheimer's pathologic change with dementia

A: Aggregated amyloid-β, T: Aggregated tau, N: Neurodegeneration
+/-: The value of a biomarker summary measure is higher/lower than the cut-point

The AT(N) framework combined with the described cut-points were used to establish the biomarker profile groups for our dataset. We then identified the sub-groups that are more susceptible to misclassification and explored their underlying causes. This is done by focusing on those groups in which the biological AD biomarkers cannot be an informative representation of the cognitive stage. For instance, individuals with normal AD biomarkers but clinical AD diagnosis are likely to be classified as non-AD class. Also, subjects with

abnormal AD biomarkers but no cognitive impairment might be identified as AD class by the model. The number of subjects in each AT(N) group was calculated for our dataset, and the probability of occurring false positive and false negative outcomes is measured, representing the contribution of biomarker shortage to the classification error.

3.4 Results and Discussion

3.4.1 Classification Results

After data preprocessing, exploratory data analysis, and feature selection, classification models (SVC, RF, and XGB) were implemented for MCI, and AD diagnosis and their performance were compared. Since the data is unbalanced, various evaluation metrics, including precision, recall, and F1-score, are reported besides accuracy. Experiments were conducted using different modalities, both separately and combined. Amyloid PET, tau PET, and MRI as single modalities, and combinations of {amyloid PET & tau PET}, and combinations of {amyloid PET & tau PET & MRI}, as multimodal scenarios were investigated, and the results are presented in Table 3.2.

Table 3.2 Classification results before feature selection for three single-modality scenarios including amyloid PET SUVRs (tracer: AV45), tau PET SUVRs (tracer: AV1451), and MRI (cortical thickness) and two multimodality scenarios including "amyloid PET SUVRs & tau PET SUVRs" and "amyloid PET SUVRs & tau PET SUVRs & MRI cortical thickness". Three machine learning models, including SVC, RF, and XGB were used, and four scores, including accuracy, precision, recall, and F1-score are reported.

			CN/MCI/AD				**CN/MCI**			
	Modality		**ACC**	**PRE**	**REC**	**F1**	**ACC**	**PRE**	**REC**	**F1**
PET	amyloid-β	**SVC**	60.2	52.6	49.7	50.4	68.9	65.2	61.6	61.9
		RF	58.6	46.4	44.5	44.5	66.9	62.4	60.1	60.3
		XGB	63.5	54.2	50.8	51.4	67.2	62.8	60.4	60.7
PET	tau	**SVC**	64.7	57.8	48.5	49.9	69.4	65.9	62.1	62.5
		RF	62.9	55.3	48.9	50.4	68.2	64.1	61	61.3
		XGB	63.1	55.8	49.3	50.9	69.2	65.5	62.8	63.2
	MRI	**SVC**	59.5	52.5	50.2	51.1	69.7	67.4	62.1	62.3
		RF	63.3	58.7	50.5	52.1	69	66.4	61.4	61.5
		XGB	62.6	57.5	50.2	52.2	65.5	61.1	58.8	58.8
& tau PET	Amyloid-β PET	**SVC**	64.2	56.2	49.9	51.3	67.8	63.7	61.3	61.7
		RF	64.9	56.5	50.7	52	71.8	69.5	64.6	65.2
		XGB	64.9	64.4	53.5	56.5	67	62.8	61.1	61.4
& tau PET	Amyloid-β PET	**SVC**	69.3	63	55.3	57.8	73.8	70.6	66.2	67.2
		RF	69	61.8	51.7	54.1	78	76.1	71.7	73
		XGB	68.8	62.9	54.6	57	78.3	78.1	70.5	72.2

			MCI/AD				CN/AD			
	Modality		**ACC**	**PRE**	**REC**	**F1**	**ACC**	**PRE**	**REC**	**F1**
PET	amyloid-β	**SVC**	74.9	66.2	64	64.8	88.6	78.8	76.3	77.4
		RF	75.9	67.6	64	65.2	89.6	81.3	76.9	78.8
		XGB	75.4	66.7	63	64.1	88.3	78.4	74.4	76.1
PET	tau	**SVC**	75.4	66.4	60.9	62	90.9	86.6	75.9	79.9
		RF	79.7	74.4	67.2	69.2	90.6	85.4	75.7	79.4
		XGB	77.5	70.4	69.2	69.7	90.6	85.4	75.7	79.4
	MRI	**SVC**	75.4	65.1	63.1	63.9	91.6	85.3	79.9	82.3
		RF	77.5	68.1	62.5	63.9	92.5	88.2	80.5	83.7
		XGB	78.2	69.5	64	65.5	90.8	83.5	77.6	80.1
& tau PET	Amyloid-β PET	**SVC**	76.5	67.8	62.4	63.7	89.9	82.7	74.9	78
		RF	78.6	71.9	65.2	67	91.5	87.3	77.6	81.3
		XGB	80.7	75.5	68.8	70.8	91.2	84.8	79.1	81.6
& tau PET	Amyloid-β PET	**SVC**	81	74.9	65.4	67.7	91.4	83.5	74.4	77.9
		RF	78.9	69.9	65.1	66.6	92.6	87.7	76.4	80.6
		XGB	78.2	68.5	61.5	63	91	81.4	75.5	78

CN: Cognitively normal, MCI: Mild cognitive impairment, AD: Alzheimer's disease

ACC: Accuracy, PRE: Precision, REC: Recall, F1: F1-score

Amyloid-β PET: SUVR values with AV45 tracer. Tau PET: SUVR values with AV1451 tracer. MRI: Cortical thickness

In terms of machine learning models, generally, SVC yields slightly less accurate scores compared to the other two models. The F1-scores of the three models for various scenarios can be seen in Figure 3.2.

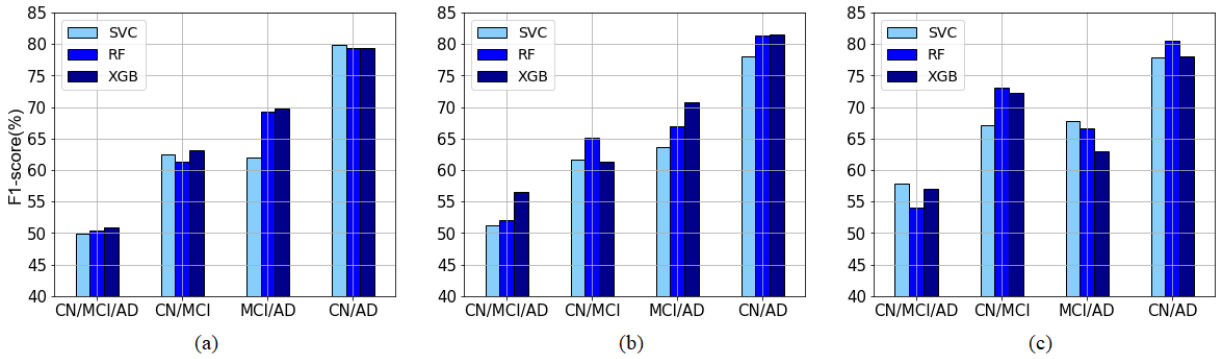


Figure 3.2 Classification F1-score before feature selection for the three machine learning models, SVC, RF, and XGB, for different classification scenarios including CN/MCI/AD, CN/MCI, MCI/AD, and CN/AD; (a) Single modality; tau PET, (b) Multimodality; tau and amyloid PET, (c) Multimodality; tau and amyloid PET and MRI.

Among single modality cases, tau PET has slightly higher scores for CN/MCI classification (early stages), and tau PET and MRI have improved results for MCI/AD and CN/AD cases. Multimodal scenarios resulted in enhanced performance in the three-class CN/MCI/AD and CN/MCI cases while not in the MCI/AD case. This is due to the fact that the feature selection has not yet been applied, and thus, in multimodal cases, the feature space is of high dimensionality, and the model could not handle it effectively. This issue is reinvestigated in the next section, where the feature selection is applied before fitting the models.

The classification scores with feature selection are shown in Table 3.3. The SVC results have improved in most cases, while the RF and XGB results have not changed significantly since these two algorithms have an embedded feature selection process and are not affected substantially by external feature selection. Figure 3.3 shows the feature selection effect on SVC and XGB F1-scores for three scenarios. In most cases, SVC with feature selection yields the highest scores, which proves the effectiveness of the proposed feature selection approach.

Table 3.3 Classification results after feature selection for three single-modality scenarios including amyloid PET SUVRs (tracer: AV45), tau PET SUVRs (tracer: AV1451), and MRI (cortical thickness) and two multimodality scenarios including "amyloid PET SUVRs & tau PET SUVRs" and "amyloid PET SUVRs & tau PET SUVRs & MRI cortical thickness". Three machine learning models, including SVC, RF, and XGB were used, and four scores, including accuracy, precision, recall, and F1-score are reported.

Modality	Classifier	CN/MCI/AD				CN/MCI			
		ACC	PRE	REC	F1	ACC	PRE	REC	F1
amyloid-β	**SVC**	62.4	57.9	52.1	53.9	69.7	66.2	62.8	63.3
	RF	61.3	52.4	49.6	50.1	68.7	64.8	61.4	61.7
	XGB	61.1	52.5	50.7	51.2	65.7	61	59.4	59.7
tau	**SVC**	65.3	55.9	53	53.9	71.9	69.6	64.7	65.4
	RF	64.9	57.9	50.4	52.1	68.7	64.8	61.9	62.3
	XGB	64.2	57	52	53.2	68.4	64.5	62.2	62.6
MRI	**SVC**	59.5	52.5	50.2	51.1	68.2	64.8	62.8	63.2
	RF	63.3	56.8	49	50.5	69.2	66.4	62.2	62.4
	XGB	62	56.8	49.2	51	68.7	65.4	62.8	63.2
Amyloid-β PET	**SVC**	67.1	61.7	54.8	56.5	73.8	73.8	65.6	66.4
	RF	64.9	59.1	51.6	53.6	72.3	70.2	65.1	65.9
	XGB	64.2	56.4	51.5	52.7	70	66.7	63.7	64.3
Amyloid-β PET	**SVC**	71.5	66.5	58.5	61.2	75.9	73.6	68.7	70.0
	RF	70.7	64.3	51.2	53.6	77.7	76.6	70.3	71.8
	XGB	69.9	62.9	55	57.3	75.6	73.1	68.5	69.7

Modality	Classifier	MCI/AD				CN/AD			
		ACC	**PRE**	**REC**	**F1**	**ACC**	**PRE**	**REC**	**F1**
amyloid-β	**SVC**	78.1	71	68.2	69.3	90.9	84.6	78.5	81.1
	RF	78.1	71.4	65.5	67.1	89.6	81.6	76	78.4
	XGB	75.9	67.7	64.7	65.7	89	80.5	73.9	76.6
tau	**SVC**	77.5	70.2	66.5	67.8	89	80.9	73.1	76.1
	RF	79.1	72.9	68.2	69.8	92.2	89.2	79.2	83.1
	XGB	75.9	68.1	66.8	67.3	89.9	83.1	75.4	78.4
MRI	**SVC**	76.4	66.8	64.8	65.6	92.1	86.3	80.8	83.2
	RF	80.3	73.3	67.4	69.3	92.7	87.6	82.4	84.7
	XGB	79.2	71.2	67.7	69	91	84.6	77.1	80.2
Amyloid-β PET	**SVC**	77	68.8	64.9	66.1	92.5	89.4	79.9	83.6
	RF	77	68.8	63.4	64.8	91.2	87.7	75.6	80
	XGB	75.9	67.1	64.1	65.1	90.6	84.4	76.1	79.4
Amyloid-β PET	**SVC**	82.4	76.6	69.5	71.8	93.3	88.9	79.4	83.2
	RF	81.7	76.9	65.9	68.4	90.6	80.5	74	76.7
	XGB	80.3	73	65	67	91.8	86.4	73.3	77.9

CN: Cognitively normal, MCI: Mild cognitive impairment, AD: Alzheimer's disease

ACC: Accuracy, PRE: Precision, REC: Recall, F1: F1-score

Amyloid-β PET: SUVR values with AV45 tracer. Tau PET: SUVR values with AV1451 tracer. MRI: Cortical thickness

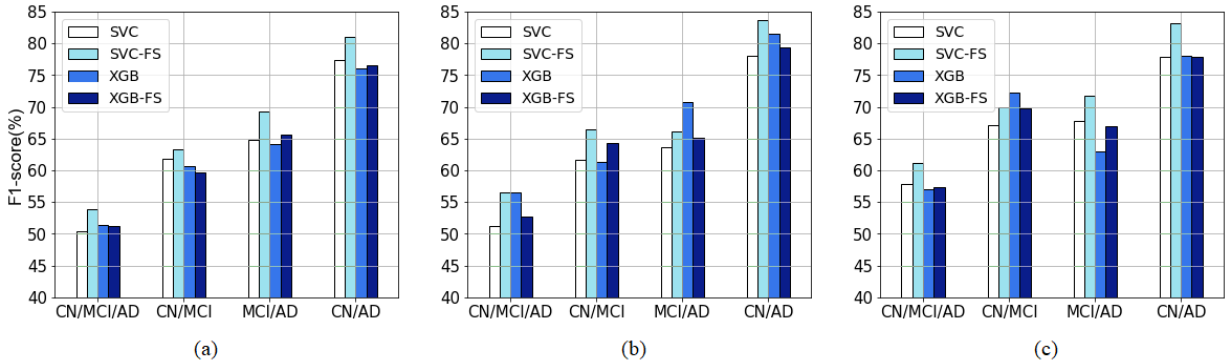


Figure 3.3 Classification F1-score before and after feature selection (FS) using two machine learning models, SVC and XGB, for different classification scenarios including CN/MCI/AD, CN/MCI, MCI/AD, and CN/AD; (a) Single modality; amyloid PET, (b) Multimodality; tau and amyloid PET, (c) Multimodality; tau and amyloid PET and MRI

Next, Figure 3.4 compares the individual modality and multimodality results. In the single modality classification, tau PET has higher scores, specifically in the CN vs. MCI case. This proves the effectiveness of tau PET compared to amyloid PET and MRI in mild cognitive impairment diagnosis, which conforms with previous studies [21]. Generally, multimodal data enhances the scores, which is more notable when feature selection is applied.

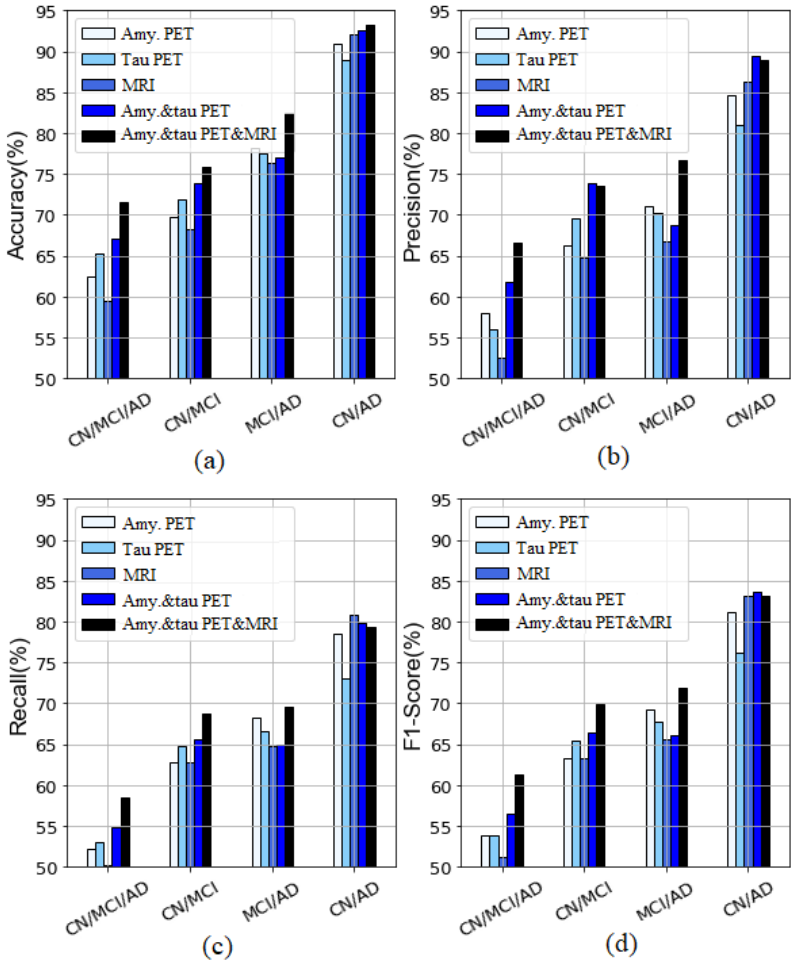


Figure 3.4 Classification scores for single-modal and multimodal scenarios after feature selection; (a) Accuracy, (b) Precision, (c) Recall, (d) F1-score

To investigate the effect of age, gender, APOE e4, and education on the classification performance, we added them to the model variables and repeated the experiments using the best-performing model and top regional features. Figure 3.5 presents the classification scores with and without the covariates age, gender, APOE4, and education. In most cases, the classification scores increased. The binary classification cases, MCI/AD and CN/AD, experienced the highest performance improvement which can be due to the higher interclass variance of covariates such as age for these classes. On the other hand, the scores for the three-class classification case, CN/MCI/AD, remained almost unchanged, which

can be due to the lower interclass variance of age between the CN and MCI classes and also the more complex nature of the multiclass classification task.

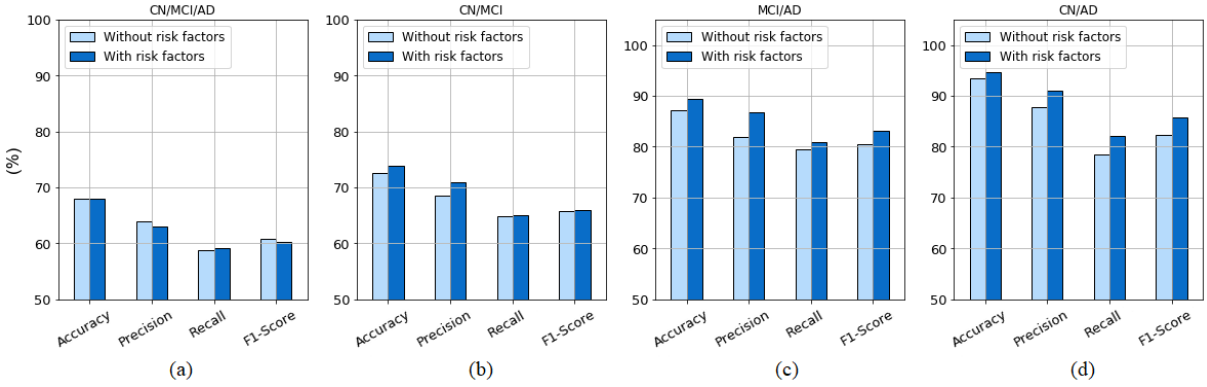


Figure 3.5 Classification scores with and without the covariates age, gender, APOE4, and education using the SVC model and top selected features for classification tasks (a) CN/MCI/AD, (b) CN/MCI, (c) MCI/AD, (d) CN/AD.

3.4.2 Biomarker Profile Grouping Results

The merit of using the National Institute on Aging—Alzheimer's Association AT(N) framework was examined to address the challenge in ascertaining discrepancies between cognitive stage (determined clinically) and biological AD (determined by the classification model using biomarkers). Biomarker profiles were thus defined based on amyloid/tau/neurodegeneration (A/T/N) positivity and negativity, as summarized in Table 3.1. The study participants were categorized according to their biomarker signature and cognitive stage. The total number of subjects falling under each category is reported in Table 3.4. The numbers are reported for two sets of cut-points; {1.11, 1.32, 2.57} and {1.11, 1.19, 2.69} for {amyloid SUVRs, tau SUVRs, and MRI cortical thickness}, respectively. The former set has a larger cut-point for tau and a smaller cut-point for MRI (confident scenario, resulting in less positive cases) compared to the second set (conservative scenario, with more positive cases). Based on this table, the inconsistencies

between the neuropathologic biomarkers and clinical diagnosis can be investigated specifically in challenging categories such as normal AD biomarkers with a dementia diagnosis and preclinical AD with cognitively unimpaired diagnosis. In the studied cohort, the "normal AD biomarker (A-T-N-) with an AD diagnosis" group includes 2 and 1 individuals based on the confident and conservative cut-points, respectively. Although this inconsistency between the biomarkers and clinical diagnosis might be partially caused by inaccurate binary biomarker grouping, it can potentially be one of the contributors to misclassification. Another controversial case is related to individuals with "preclinical Alzheimer's disease biomarkers" (A+T+N- and A+T+N+). As seen in Table 3.4, this group has a considerable number of subjects in all three cognitive stages making the classification task even more challenging.

Table 3.4 Grouping the study participants into AT(N) biomarkers categories and their corresponding clinically diagnosed cognitive stage (CN, MCI, and AD). The AT(N) groups are defined using two different cut-points for each biomarker. Confident cut-points {1.11, 1.32, 2.57} and conservative cut-points {1.11, 1.19, 2.69} were used for amyloid SUVRs, tau SUVRs, and MRI cortical thickness, respectively. The distribution of subjects shows that in each biomarker profile specifically for the preclinical AD group (A+T+N- and A+T+N+), subjects can belong to any of the three cognitive stages, which is due to the heterogeneity of the disease. This results in a more challenging classification of the cognitive stage. For the confident cut-points, more subjects are categorized in the A-T-N- and A+T-N- groups, while for the conservative cut-points, groups with more positive biomarkers include a larger number of subjects. This is expected as the confident cut-point case has a larger threshold for tau SUVR and a smaller threshold for cortical thickness compared to the conservative cut-point case.

	Clinically diagnosed cognitive stage					
	Confident cut-points			**Conservative cut-points**		
	CN	**MCI**	**AD**	**CN**	**MCI**	**AD**
A- T- N-	82	38	2	56	23	1
A+ T- N-	41	15	5	23	9	2
A+ T+ N- **A+ T+ N+**	9	14	12	22	21	16
A+ T- N+	2	3	2	7	2	1
A- T+ N- **A- T- N+** **A- T+ N+**	4	9	0	30	24	1

CN: Cognitively normal, MCI: Mild cognitive impairment, AD: Alzheimer's disease
A: Aggregated amyloid-β, T: Aggregated tau, N: Neurodegeneration

Table 3.5 Grouping the study participants into AT(N) biomarkers categories and their corresponding clinical and predicted cognitive stage (CN, MCI, and AD). The AT(N) groups are defined using confident cut-points {1.11, 1.32, 2.57} for amyloid SUVRs, tau SUVRs, and MRI cortical thickness, respectively. For the normal biomarker profile (A-T-N-), more subjects were predicted as the CN class (compared to the clinical diagnosis) due to the dominance of CN subjects in this specific AT(N) group. The Alzheimer's pathological change group (A+T-N-) experienced a similar but less severe situation than the previous group. In the preclinical AD group (A+T+N- and A+T+N+), all three cognitive classes include a significant portion of subjects for both clinical and predicted cases.

	Clinical cognitive stage			**Predicted cognitive stage**		
	CN	**MCI**	**AD**	**CN**	**MCI**	**AD**
A- T- N-	137	52	4	160	33	0
A+ T- N-	66	20	8	71	17	6
A+ T+ N- **A+ T+ N+**	13	24	18	15	25	15
A+ T- N+	2	4	3	3	2	4
A- T+ N- **A- T- N+** **A- T+ N+**	5	9	0	6	8	0

CN: Cognitively normal, MCI: Mild cognitive impairment, AD: Alzheimer's disease
A: Aggregated amyloid-β, T: Aggregated tau, N: Neurodegeneration

To further investigate this scenario, we reconstructed the AT(N) biomarker-cognition table for the predicted cognitive stage aside from the clinically diagnosed cognitive stage. Table 3.5 represents the results for the clinical and predicted diagnosis side by side. It should be noted that here we used a different case study than Table 3.4. As can be seen from the results, for the normal biomarker group (A-T-N-), all dementia subjects and some of the MCI subjects were misclassified as the CN group (false negative). A less severe outcome is seen for the AD pathological change group (A+T-N-), where some AD and

MCI subjects were misclassified as CN. As for the challenging preclinical AD group (A+T+N- and A+T+N+), a clear conclusion cannot be drawn solely from Table 3.5. Thus, a classification confusion matrix was constructed for the specific case of preclinical AD, as shown in Table 3.6. From this table, it is clear that many CN subjects were misclassified as MCI, and a large number of AD subjects were misclassified as MCI.

Table 3.6 Classification confusion matrix for the AT(N) preclinical AD group (biomarker profiles A+T+N- and A+T+N+). For the CN class (true label), a significant portion of subjects (6 out of 13) was classified (predicted label) as MCI and AD, which can be related to those preclinical AD individuals that have not yet advanced to AD. On the other hand, a considerable number of AD subjects (true label) were classified (predicted label) as MCI and CN, which could belong to those AD subtypes with a different pattern and less severe biomarker levels. Overall, the classification scores for this preclinical AD category are: accuracy=56.4% , precision=57.3% , recall=56.4% , f1-score=55.5% .

True/Pred	**CN**	**MCI**	**AD**
CN	**7**	4	2
MCI	7	**14**	3
AD	1	7	**10**

As described in this study, one challenge in the classification problems is biomarker insufficiency. This may result in a disconnection between biomarkers and clinical diagnosis to some extent. Studies revealed that almost 30% of clinically unimpaired elderly participants have AD in postmortem examinations or have abnormal amyloid deposition [42], [74]. In our study, in one of the scenarios (Table 6), 6.5%-16% (9-22 individuals) of the CN group have preclinical AD with abnormal amyloid and tau pathology for the two cut-point levels, as seen in Table 6. It is anticipated that the classification model classifies some of these individuals as MCI or AD groups since both AD-specific biomarkers

(amyloid and tau) are abnormal in this case (false positive). This was confirmed in Table 8, where almost half of the CN subjects were misclassified as MCI and AD. Moreover, for the same preclinical AD group, a large number of AD subjects were misclassified. This can be explained by the heterogeneity of AD, where some AD subjects with less severe biomarkers are predicted by the model as non-AD and vice versa. The results proved the preclinical AD subjects to be one of the most challenging groups for the model, with a classification accuracy of 56%, which is lower than the overall accuracy of 65% for all subjects of the scenario presented in Table 7. These outcomes were expected since the preclinical biomarker profile includes subjects in all three cognitive stages which is due to the heterogeneity of the disease and the lack of sufficient biomarkers required for a more accurate delineation of the classes. Similarly, the "normal AD biomarker" (A-T-N-) and "non-Alzheimer's pathologic change" (A-) groups are also susceptible to misclassification as they have non-AD-specific biomarkers, but some are labeled as MCI (AD prodromal stage) and AD in the ADNI dataset. It has been shown in other studies that 10% to 30% of clinically diagnosed AD cases do not have AD at autopsy or have normal AD biomarkers [42], [74]. In the ADNI cohort used in our study, 10-20% of subjects were detected with the described condition. In the classification process, the normal biomarkers are likely to predict a cognitively normal stage rather than AD (false negative). These results can be explained by the fact that the clinical diagnosis and cognitive labeling practices are generally based on symptoms and are independent of the biomarkers. The outcomes reveal the insufficiency of the available biomarkers in making an accurate prediction of the clinically defined cognitive stage.

4. TRANSFER LEARNING

4.1 Introduction

In this chapter, a transfer learning framework is presented for building classification models for a target dataset with a limited amount of labeled data. For this purpose, a dataset with a sufficient amount of data is used as the source for transferring knowledge. The proposed approach uses the boosting mechanism to penalize source instances with different distributions than the target data. The framework is built on top of the GBM algorithm, and the residuals of the GBM base learners are used for defining weights for the source and target instances. A detailed description of the proposed algorithm is presented. To evaluate the effectiveness of the model, a set of experiments are conducted. The model performance is compared to two baseline scenarios where a model is trained solely based on either the source data or the target data. The results support the satisfactory performance of our boosting-based transfer learning model for multimodal multiclass classification in a source and target domain setup with different distribution of feature and label spaces.

4.2 Data domains for transfer learning

In the transfer learning scenario, if the target domain does not include any label information, it is known as transductive learning, while if labeled data is available, the problem is referred to as inductive transfer learning [43]. Transfer learning approaches can also be explained based on the degree of consistency between the target and source domains. Let X_S and X_T be the feature space of the source and target domains, and Y_S and Y_T be the label spaces, respectively. The subscripts S and T refer to the source and target domains. The data of the two domains may vary in different aspects. When the feature

spaces or the label spaces are different ($X_S \neq X_T$, $Y_S \neq Y_T$), the problem is known as heterogeneous transfer learning. In contrast, homogeneous transfer learning is the case when the source and target domains have the same features and labels [26]. For instance, in the AD classification, if different regional features or different modalities are used for the source and target domains, we are dealing with a heterogeneous transfer learning problem. Alternatively, the feature spaces could be the same, but their marginal distribution can be different ($P(X_S) \neq P(X_T)$). Some studies only try to mitigate this discrepancy between the marginal distributions, which is known as sample selection bias [75], [76]. As an example, a medical dataset may only include completely healthy subjects, while another dataset might have a minimum requirement of cognitive concern even for the cognitively normal group. Similarly, for the AD group, different datasets may differ in the severity of the biomarkers and the disease progression. These situations can lead to a shift in the distribution of the input data.

The other transfer learning scenario is when the source and target domains have different conditional probability distributions ($P(Y_S|X_S) \neq P(Y_T|X_T)$). This condition is known as context feature bias. In this case, given identical input features, the target variable differs for the two domains. In the AD context, subjects with similar biomarkers may be mapped to different diagnosis groups due to various reasons. This can be partially due to the subjective procedure of data annotation and also heterogeneity of the disease, which can be explained by the AD subtypes and the risk and protective factors. For instance, in an AD group, younger subjects tend to have different levels and patterns of biomarkers than older ones, and if age is not considered, those subjects might be mapped to the wrong cognitive group.

Besides the marginal and conditional probability distributions of the input X, there can be a mismatch between the class space for the source and target domains ($Y_S \neq Y_T$). The label space of the two domains may vary in terms of the number of classes and the class labels themselves. In our study, the source domain might be dealing with a binary classification between the CN and AD groups, while the target domain may focus on a multiclass classification or a binary classification with different labels such as early MCI and late MCI. The other scenario is when the data is unbalanced between the two sources resulting in a disparity between the probability distribution of the labels ($P(Y_S) \neq P(Y_T)$). In the ADNI dataset, there is a lower number of AD subjects compared to CN and MCI, while the Mount Sinai dataset has balanced AD and CN but more MCI subjects. In this study, the feature space of the two datasets is the same, while their marginal and conditional distributions can be different. As for the label space, there is a mismatch between the class labels and their distributions in some scenarios.

4.3 Data

4.3.1 Participants

For our analysis, the clinical data were obtained from two datasets. For the target domain, the participants are part of the 1Florida Alzheimer's Disease Research Center (ADRC) in a 5-year study since 2015 at the Mount Sinai Medical Center. The number of subjects for each group of CN, MCI, and AD in this dataset is 53, 141, and 45, respectively. As part of the 1Florida ADRC baseline analysis, a wide range of neuropsychological and clinical tests were performed, as well as neuroimaging studies such as structural MRI and PET/CT scans to measure fibrillar amyloid plaques. Each participant underwent clinical

assessments, including the Clinical Dementia Rating (CDR) and the Mini Mental State Examination (MMSE). The neuropsychological examination also incorporates the Hopkins Verbal Learning Test-Revised (HVLT-R). For each participant, structural MRI was acquired at Mount Sinai Medical Center using a Siemens Skyra 3T MRI scanner. The FreeSurfer software was used for brain segmentation using a 3D T1-weighted sequence (MPRAGE) with isotropic resolution of 1.0 mm. A 3D Hoffmann brain phantom was used for PET scan acquisition. PET tracer [18-F] florbetaben 300 MBQ was infused 70-90 minutes before scanning. Each subject was scanned on a Siemens Biograph 16 PET/CT scanner (55 slices/frame, 3 mm slice thickness, 128 × 128 matrix). The scans were then transformed to a 128 × 128 × 63 dimension with the size of 0.21 × 0.21 × 0.24 cm.

On the other hand, the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu) was used as the source domain. ADNI was launched in 2003 as a public-private partnership, directed by Principal Investigator Michael W. Weiner, MD. The primary objective of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. For up-to-date information, see http://www.adni-info.org. The participants' demographics and Mini-Mental State Examination (MMSE) scores for ADNI 3 cohort used in this study are given in Table 4.1. The modalities MRI and amyloid PET (agent: 18F-AV45) were included in the analysis. The T1 weighted MRI scans have gone through preprocessing, gradient wrapping, scaling, shading artifact, and inhomogeneity corrections. For skull stripping and cortical and subcortical segmentation of the T1 images, the FreeSurfer package was used. The segmented MRI scans were then co-registered with the florbetapir scans to measure the volume-weighted average of

amyloid deposition in regions of interest and compute the standardized uptake value ratio (SUVR). Besides these neuroimaging biomarkers, AD risk and protective factors, including age, gender, APOE4, and education, are also used as independent variables for each subject. These variables, when combined with the biomarkers, can help with the disease heterogeneity challenge.

Table 4.1 Participant demographics and mini-mental state examination (MMSE) score for different diagnosis groups of the ADNI3 cohort.

Groups	Subject (f/m)	Age (year)	Education (year)	MMSE
CN	277(153/124)	71.80±5.70	16.67±2.47	28.63±2.12
MCI	378(155/223)	71.26±7.66	16.25±2.61	26.87±4.20
AD	67(26/41)	73.41±8.78	16.43±2.35	22.37±2.39

4.3.2 MRI artifact restoration

One of the important preprocessing steps for the Mount Sinai dataset was MRI shading artifact correction. Shading artifact occurs due to low signal intensity in some parts of the image. Some causes of signal intensity artifact include non-uniform excitation in some of the nuclei, improper coil positioning and tuning, and magnetic field inhomogeneity. Artifacts can affect the interpretation and diagnostic quality and should be addressed before further analysis [77]. To compensate for the shading artifact, bias field correction is implemented. Also, the effect of artifact correction on the Grey Matter (GM) and White Matter (WM) is studied.

4.3.2.1 Bias field correction

Two experiments have been conducted using the Statistical Parametric Mapping (SPM) tool. First, the whole image went through the SPM bias field correction. For a sample T1 MRI scan, Figure 4.1 shows the original image, bias field, SPM corrected image, and the corrected regions. The corrected regions are the voxel-wise subtraction of the original image and the corrected one. As seen, the most affected regions are the top and bottom of the image. Moreover, some other parts of the image are also affected. The effect of these changes on the GM and WM volumes will be discussed in the next section.

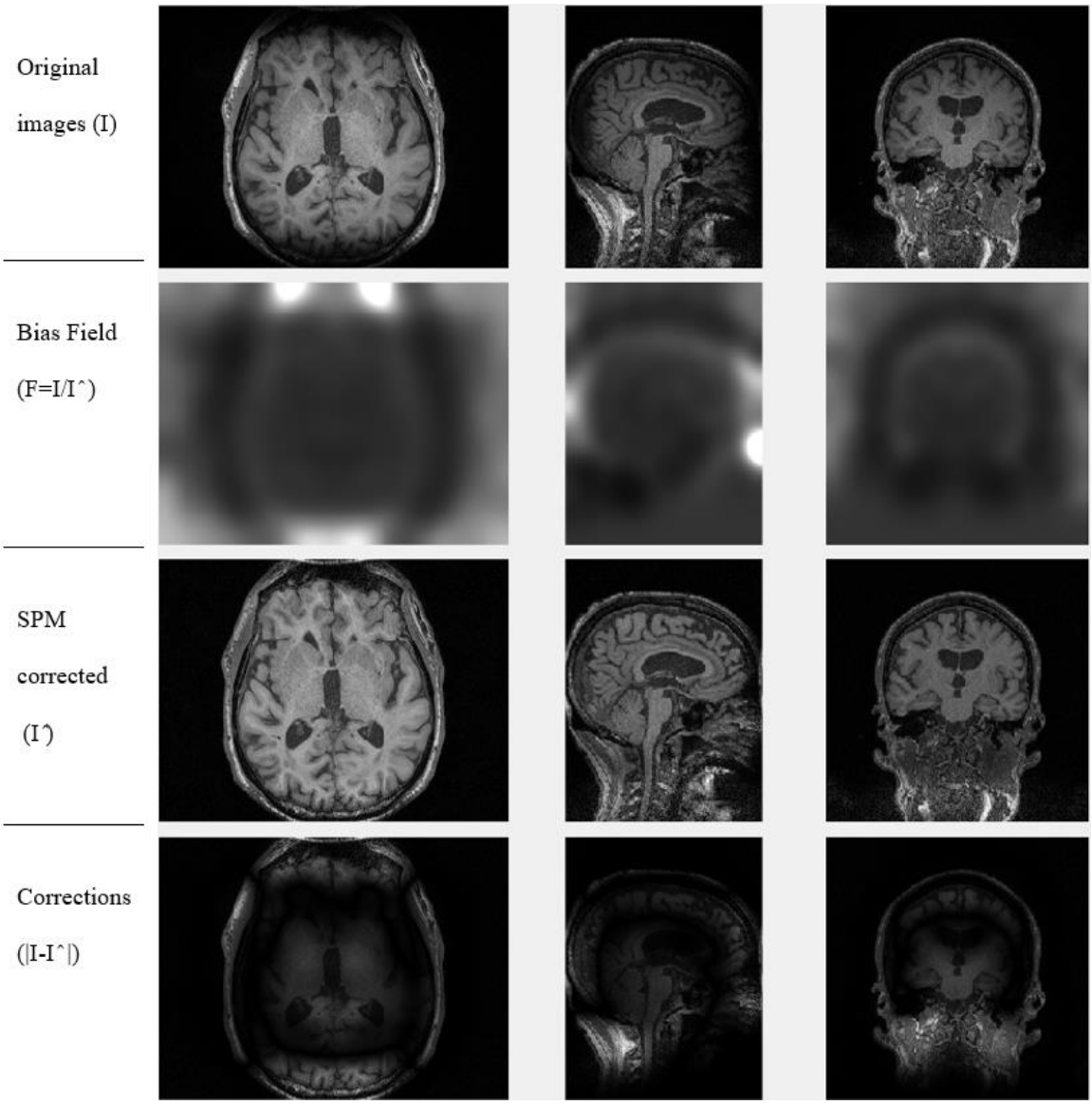


Figure 4.1 Bias field correction using SPM

The bias field correction can be limited to only the top and bottom portions of the image since the concerned shading artifact is mainly in those regions. For this purpose, the bias field and correction images from the previous step can be used. Using an edge detection operator, the desired affected regions (top and bottom) are filtered to create mask.

Then, using a Gaussian filter with a standard deviation of σ=8, the edges are smoothened, as shown in the figure. The process of creating the mask is shown in Figure 4.2.

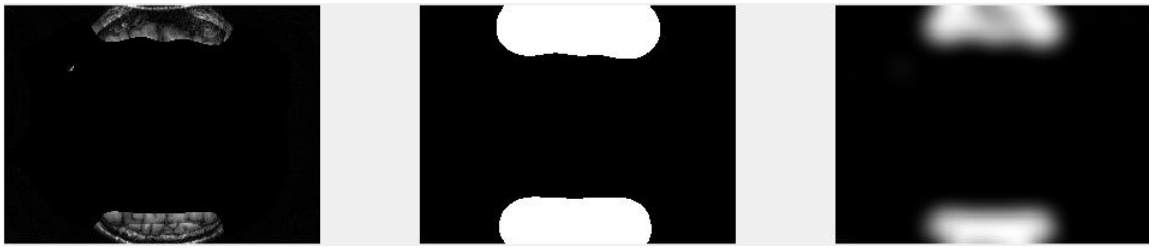


Figure 4.2 Creating a Gaussian mask for top and bottom shading correction

Next, the gaussian mask was used to implement bias field correction on specific regions of the image. The resulting corrected image and also the corrected regions (voxel-wise subtraction of the original image and the corrected one) are illustrated in Figure 4.3. As seen, the image is just corrected in the top and bottom regions, and other parts are intact.

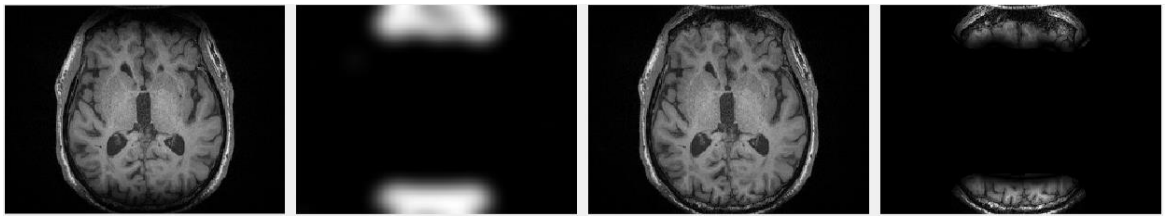


Figure 4.3 Bias field correction using Gaussian mask

4.3.2.2 Segmentation

In this section, the effect of bias field correction on GM and WM segmentation is investigated. Figure 4.4 presents the GM, WM, and CSF segmentation of the original images (before correction) using SPM.

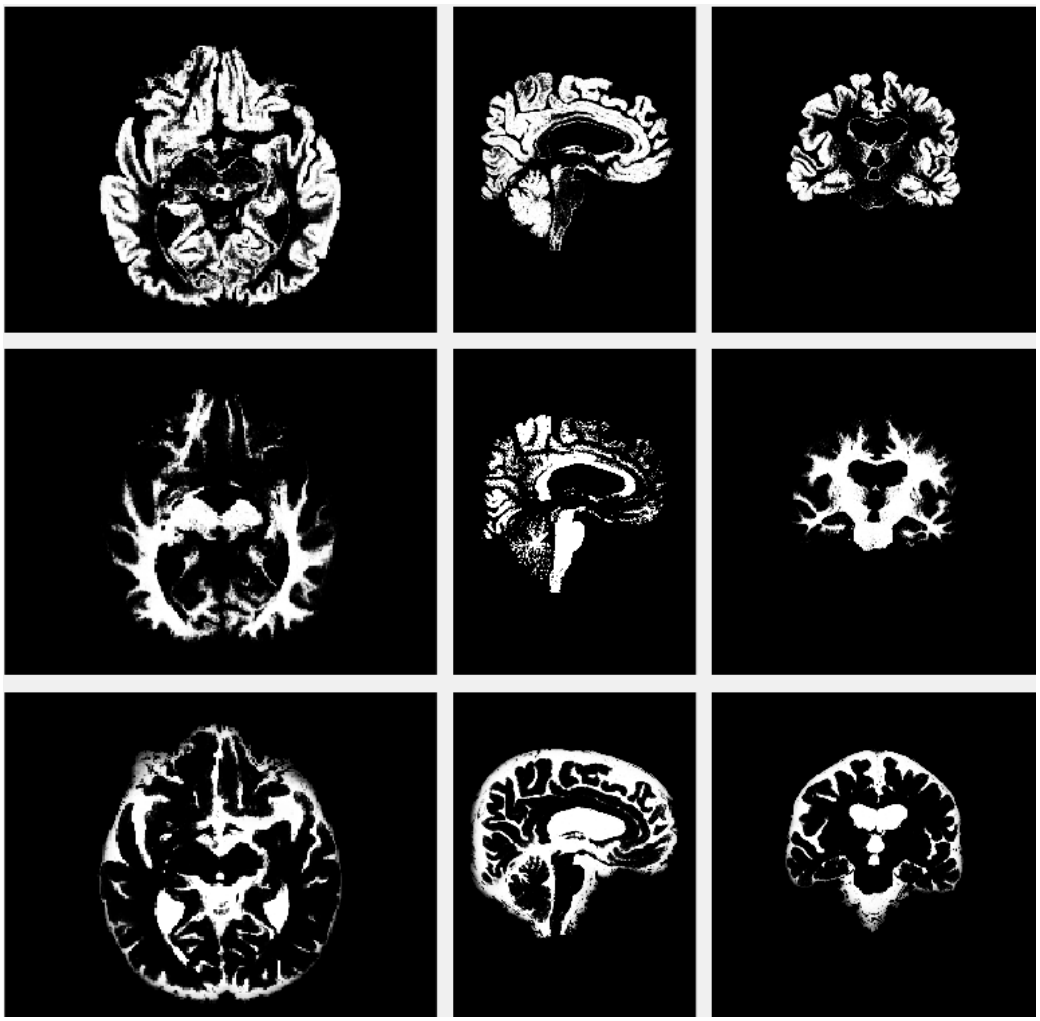


Figure 4.4 Gray matter (top), White matter (middle), and Cerebrospinal fluid (bottom) segmentation using SPM.

Next, the segmentation is accomplished for the corrected images with and without a mask, and the results are shown in Figure 4.5. The most noticeable correction can be seen in region C (bottom), where the GM and WM are restored in the corrected images compared to the original image. Moreover, some other differences between the original and corrected images can be identified as marked in regions A and B. For example, in region A, the GM volume is lower, and the WM volume is higher in the corrected image without a mask (Figure 4.5 (b)) compared to the original one (Figure 4.5 (a)). In contrast,

in the same region A, the GM volume in the corrected image with mask (Figure 4.5 (c)) is the same as the original image since, in this approach, that region of the image was intentionally left untouched.

(a) (b) (c)

Figure 4.5 GM and WM segmentation for (a) the original image, (b) corrected without mask, and (c) corrected with mask.

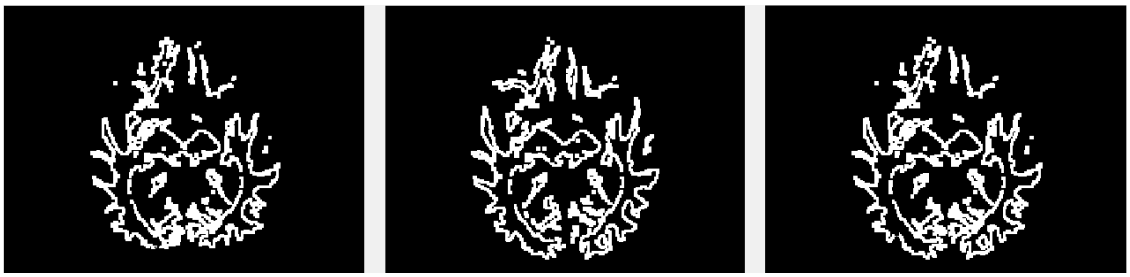
However, for region B, the GM volume is higher for the corrected image (b) than the original image (a). The ratio of WM/GM volume for the original and corrected images is 0.8060, 0.8232, and 0.8300, respectively.

In sum, if shading artifact occurs in specific regions, the bias field correction can only be implemented in those regions. The effectiveness of this correction can be investigated using the WM and GM segmentation performance and the WM/GM contrast. The contrast can be calculated as follow:

$$Contrast = \frac{(\text{mIWM} - \text{mIGM})}{(\text{mIWM} + \text{mIGM})} \times 100 \qquad (4\text{–}1)$$

where mIWM and mIGM are the mean intensity of white matter and gray matter. For a sample MRI slice, the contrast for the original image and the first and second corrected scenarios is 0.4260, 0.4524, and 0.4400, respectively.

The contrast alone cannot be a comprehensive measure of the GM/WM segmentation performance. This is because the bias field correction not only changes the voxels intensities but also changes the volumes and GM/WM borders. Thus, although the contrast of the SPM corrected MRI is the highest, it cannot be concluded that this approach provides the best segmentation performance. One solution is that the segmentation results are visually investigated by experts. The other point is that the defined GM/WM contrast will provide a single measure for the whole image. However, after the correction, the image contrast may be improved in some regions but deteriorated in other regions. Thus, it may be beneficial to compute the contrast in small patches of the image. For this purpose, a 3×3 contrast operator is defined and convolved with the image. The resulting matrix can be called the contrast matrix. Figure 4.6 illustrates the contrast matrix of the original and the corrected images.



(a) (b) (c)

Figure 4.6 The contrast matrix for (a) the original image, (b) the corrected image without mask, and (c) the corrected image with mask

The relative contrast (subtraction) of the three scenarios is calculated and represented in Figure 4.7. As seen, the SPM is affecting the contrast all over the image, while the SPMM is just focusing on the top/bottom regions.

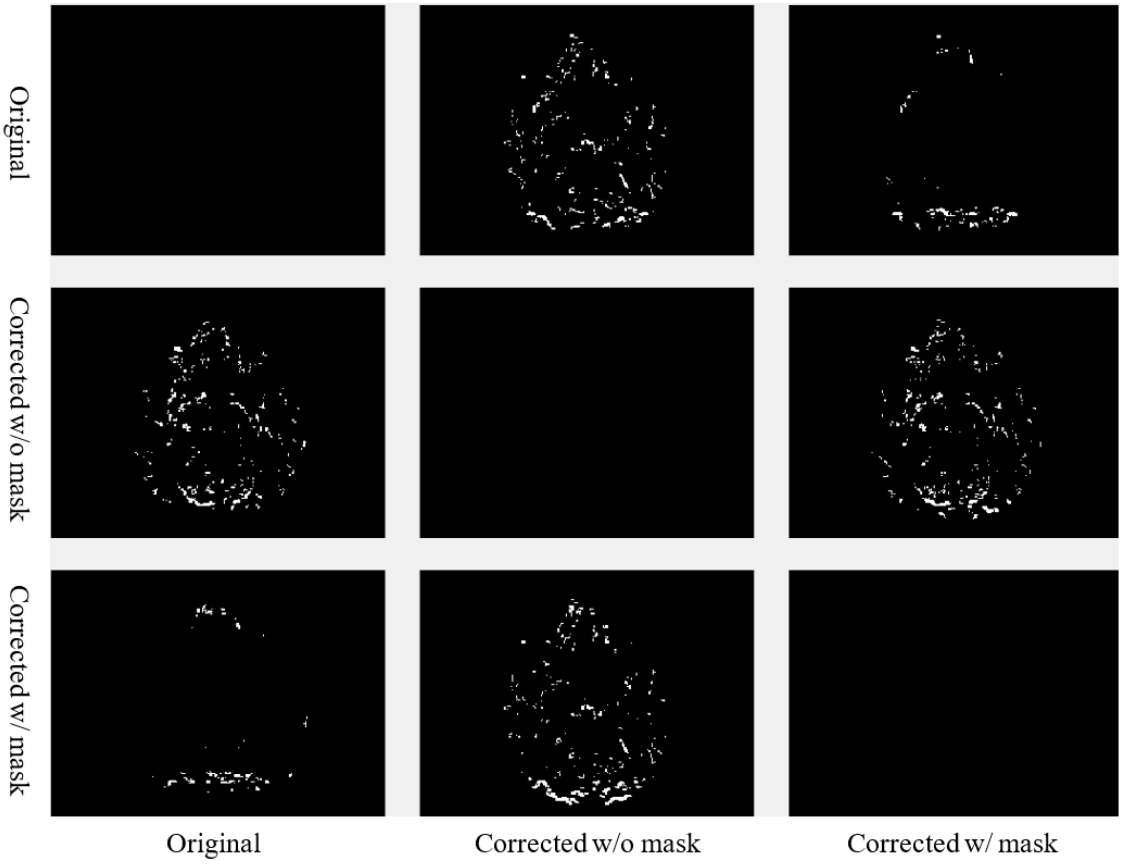
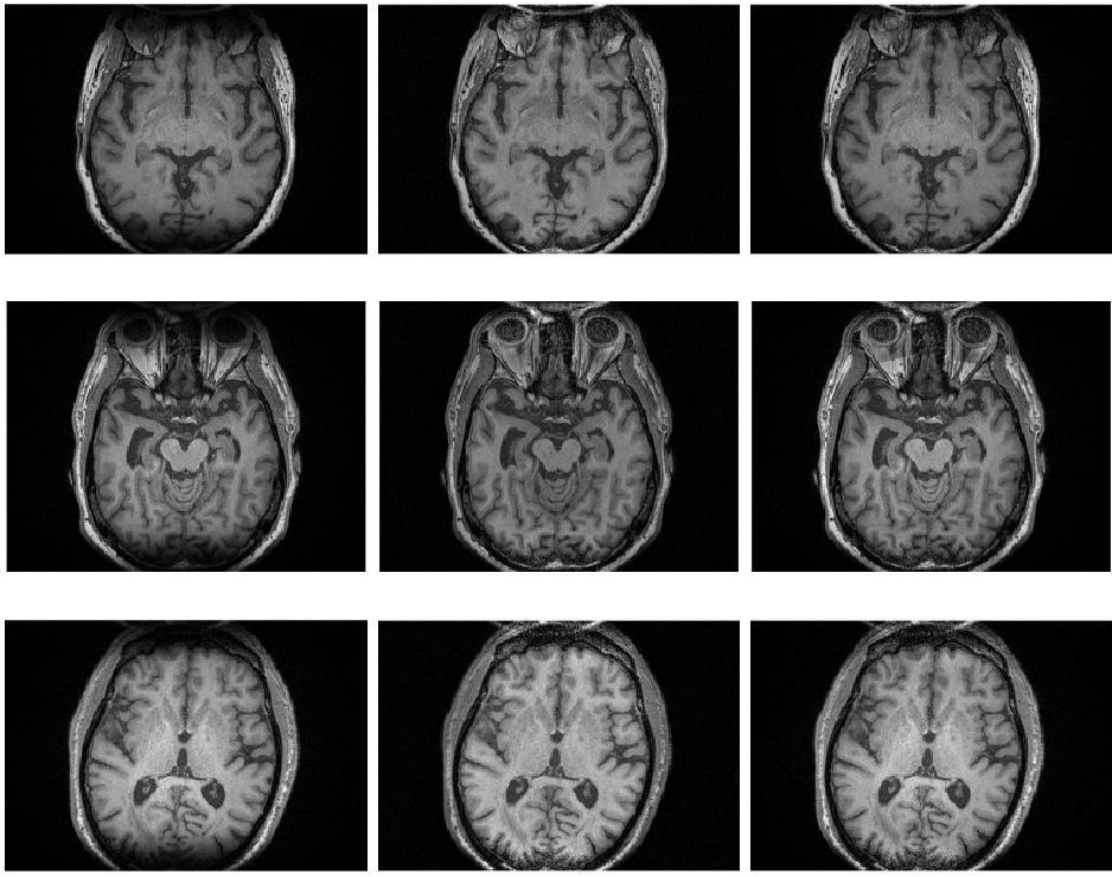


Figure 4.7 Relative contrast for the original and corrected images

Finally, the resulting corrected MRI and the original MRI are shown in Figure 4.8 for one MRI slice of three different subjects.



(a) (b) (c)

Figure 4.8 The bias field correction results for three subjects, (a) the original image, (b) the corrected image without mask, and (c) the corrected image with mask

4.4 Method: The proposed transfer learning approach

A source domain can contribute to a target domain learner directly through transferring the data. This category of transfer learning is known as instance-based transfer, which is the base of the approach of this study. Initially, the source data is added to the training set since the labeled target data is usually not enough to train an effective model. The main idea is to take advantage of the same-distribution portion of the source data and gradually fade the effect of the misleading part with a different distribution. To do so, we present an extended version of GBM for transfer learning (TrGB). GBM, which was first introduced in [78] by J. H. Friedman, is a boosting-based learning algorithm that combines

a sequence of base learners. Each base learner focuses on the error (residuals) of the previous learner, and this process is repeated until the error is less than a predefined threshold. A final prediction is made based on the combination of the response of all learners. The overall model variance is low because of using simple base learners, and a low bias is achieved through the boosting and ensembling mechanism. To extend the GBM idea to transfer learning, an instance weighting mechanism is added to penalize those source domain instances with a different distribution than the target instances. The instance weights are a function of the power of a base learner and the residual of the instance prediction using the model of the previous step. If an instance of the source domain is misclassified, based on its residual, a weight smaller than 1 is assigned to it so that its effect on the next iteration will reduce. The structure of the presented transfer learning approach is summarized in Figure 4.9.

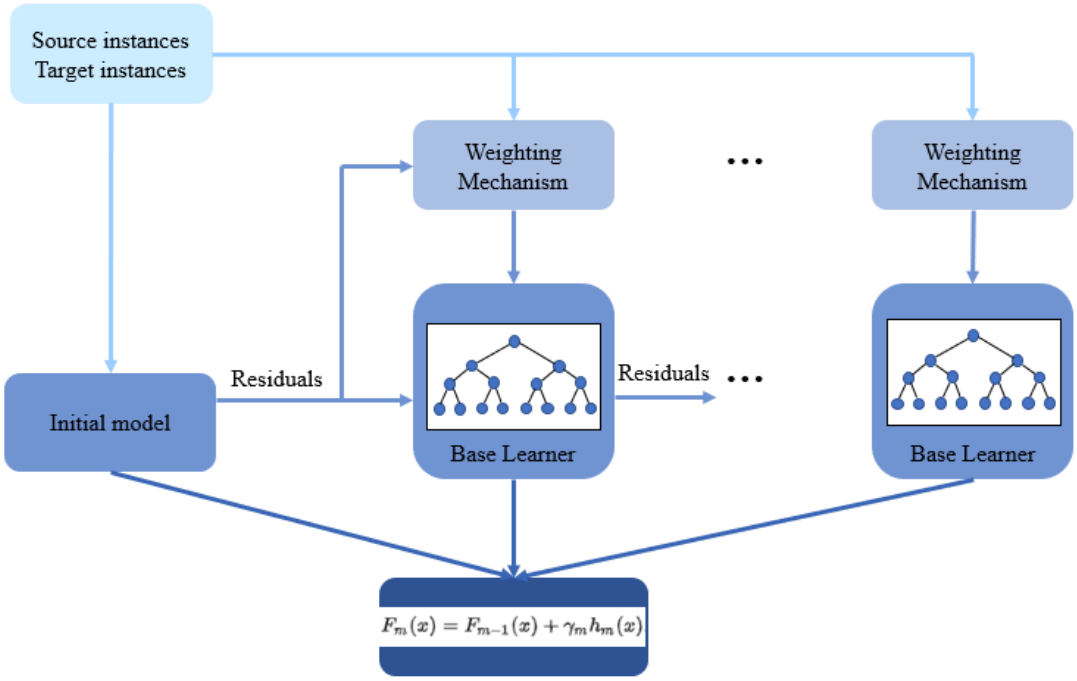


Figure 4.9 The structure of the proposed gradient boosting-based transfer learning approach (TrGB)

Algorithm1: TrGB

Input: Labeled source data: $\{(x_i, y_i)\}_{i=1}^{m}$, and labeled target data $\{(x_i, y_i)\}_{i=m+1}^{n}$.
A differentiable loss function $L(y_i, F(x))$.
The number of iterations T.

a. Initialize the model as $F_0(\mathrm{x}) = \arg min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$

b. For t=1 to T:

1. Calculate the pseudo-residuals:
$r_{it} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x_i)=F_{t-1}(x_i)}$ for i=1, …, n.

2. Calculate the performance of each weak learner (λ) based on the residuals of target domain data:
$[r_{t-avg}]_{i=m+1}^{n}$
$\lambda = \frac{1}{2}\log\left(\frac{1-r_{t-avg}}{r_{t-avg}}\right)$

3. Calculate weights for both source and target domain instances:
$w_{it} = 1 \qquad if\ (\frac{1}{2} - r_{it})\lambda > 0$

$$w_{it} = \begin{cases} e^{-2\left(r_{it}-\frac{1}{2}\right)\lambda} & \text{for i} = 1, \ldots, \text{m} \\ e^{2\left(r_{it}-\frac{1}{2}\right)\lambda} & \text{for i} = \text{m}+1, \ldots, \text{n} \end{cases} \qquad if\ (\tfrac{1}{2} - r_{it})\lambda < 0$$

4. Fit a base learner (e.g., decision tree) to pseudo-residuals using weighted source and target domain instances $\{w_{it}x_i, r_{it}\}_{i=1}^{n}$ i = 1, …, m (source); i = m+1, …, n (target).

5. Optimizing the following equation and determining the corresponding γ_m:

$$\gamma_m = \arg min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(\mathrm{x}_i) + \gamma h_m(x_i))$$

6. Updating the model:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

The framework of the proposed TrGB model is presented in Algorithm 1. The algorithm is built on top of the original GBM structure with proper modifications to enable transfer learning. As seen, the input data includes the source data (i = 1, …, m) and the labeled part of the target data (i = m+1, …, n). Since the loss gradient plays a crucial role in GBM, a differentiable loss function is required. For the binary classification problem, we used negative log-likelihood as the loss function as follows

$$-\sum_{i=1}^{n} y_i \log(p) + (1 - y_i)\log(1 - p) \tag{4-2}$$

where y_i is the actual class label for the i^{th} instance and p is the predicted probability. In step a, the model is initialized with a γ value that minimizes the loss function. To do so, the derivative of the loss function is calculated, which is equal to the residuals ($p-y_i$) for the log-likelihood loss function. Thus, for all instances, the initial model would be the mean value of y_i in terms of probabilities and $0.5\log((1+y_{avg})/(1-y_{avg}))$ in terms of log(odds) as shown in the original GBM algorithm [78]. In the next step, for the first decision tree (t=1), pseudo-residuals (r_{it}) are calculated for both source and target instances using the gradient of the loss function as follows:

$$r_{it} = -\left[\frac{\partial L(y_i,\ F(x_i))}{\partial F(x_i)}\right]_{F(x_i)=F_{t-1}(x_i)} \tag{4-3}$$

where r_{it} is the pseudo-residuals for the i^{th} instance and t^{th} iteration tree. For the loss function used in this study, pseudo-residuals are simply the difference between the predicted probability and the true values (y_i-p). Next, the effectiveness of the t^{th} learner denoted by λ is determined as follows

$$\lambda = \frac{1}{2}\log\left(\frac{1 - r_{t-avg}}{r_{t-avg}}\right) \tag{4-4}$$

where $r_{t\text{-}avg}$ is the average of the absolute value of residuals over the target data instances. Since λ is defined as a measure of the goodness of a learner, the source domain instances are not included in it. This equation maps the value of $r_{t\text{-}avg}$, which is between [0, 1], into values between [-∞, +∞]. Lower values of $r_{t\text{-}avg}$ close to 0 are associated with large negative values of λ, and values of $r_{t\text{-}avg}$ close to 1, correspond to larger positive values of λ. Thus, a good model will have low $r_{t\text{-}avg}$ and high positive λ.

Using the λ parameter and the residuals, weight coefficients are calculated and assigned to the instances. For a positive value of λ (powerful model) and an instance with a small residual (correctly classified), the weight of the instance will not be changed for the next iteration. However, for a positive value of λ, and a misclassified instance, a different weight is assigned to the instance depending on the fact that the data is from the source or target domain. If the misclassified instance comes from the target dataset, a large weight is given to the instance in order for the model to strengthen its impact on the model for the next iteration. On the other hand, a misclassified instance from the source data is given a lower weight to have a more negligible effect on the model. The weights equation is given below.

$$w_{it} = \begin{cases} e^{-2\left(r_{it}-\frac{1}{2}\right)\lambda} & \text{for i} = 1, \dots, \text{m} \\ e^{2\left(r_{it}-\frac{1}{2}\right)\lambda} & \text{for i} = \text{m}+1, \dots, \text{n} \end{cases} \qquad if\ (\frac{1}{2} - r_{it})\lambda < 0 \tag{4-5}$$

where w_{it} is the weight of the i^{th} instance in the t^{th} iteration. The reasoning for such a weighting mechanism is that a misclassified instance from the source domain is likely to have a different distribution than the target data, and its effect needs to be minimized. The equations in step b.3 of the algorithm implement the described weighting procedure.

The next steps are similar to the original GBM algorithm. First, a base regression learner such as a decision tree is trained using weighted source and target instances as the input data and pseudo-residuals as the target variable. This learner will try to fix the errors (residuals) of the previous tree by minimizing the loss function, as shown in step b.5 of the algorithm. Finally, the model is updated using the new predictions of the regression tree (step b.6). The whole process will be repeated for the next value of t until it reaches the maximum value of iterations or until the loss is lower than a predefined threshold.

4.5 Results and Discussion

To evaluate the effectiveness of our transfer learning framework, a set of experiments was conducted on the Mount Sinai data as the target domain and ADNI as the source domain datasets. The regional cortical thickness from MRI and SUVR values from amyloid PET was used as the feature set for our predictive models. Two single modality scenarios for MRI and amyloid PET and one multi-modality scenario for the combination of MRI and PET have been tested. As for the clinical diagnosis groups, the Mount Sinai dataset includes Normal, PreMCI Clinical, PreMCI Neuropsychology, non-amnestic MCI, Early MCI, Late MCI, and Dementia. In this study, four class labels including Normal, Early MCI, Late MCI, and Dementia from the Mount Sinai dataset and three class labels including Cognitively Normal, MCI, and AD from the ADNI dataset were used. For the label space, two main scenarios were investigated. First, the same label spaces were used for both source and target domains. The purpose of this test was to use the knowledge in the source model for the same classification problem (e.g., CN/MCI, or CN/AD) in the target domain. In the second step, the knowledge was transferred between different label spaces. More specifically, the information from the CN vs. AD classification task from the source domain (ADNI) was used for the Early MCI vs. Late MCI classification in the target domain. As explained, this is one of the scenarios of the transfer learning that is useful here as the amount of labeled data for the Early MCI, and Late MCI groups is too limited for building a model from scratch. In order to report the classification metrics, a 5-fold cross-validation was implemented. Since the datasets are imbalanced, besides accuracy, other evaluation metrics, including Precision, Recall, and F1-score, are also calculated and reported.

For comparison purposes, a baseline case is defined where no information from the source data is used, and a model is learned solely based on the labeled data from the target domain. In another case, we assumed that there are no labeled instances from the target domain, and a model is trained on the source data and tested on the target data. The final scenario uses the proposed TrGB approach, which is expected to outperform the other two cases. As a first step, an experiment is conducted on the ADNI dataset without making use of transfer learning. The classification scores for the three tasks of binary and multiclass classification using multimodal data are presented in Table 4.2. It should be noted that the target data is not included in this experiment. As expected, the most challenging tasks remain the multiclass classification and the binary classification of the CN/MCI cases.

Table 4.2 Classification scores for three classification tasks using the ADNI dataset

	CN-AD	CN-MCI	CN-MCI-AD
Accuracy	94.5	65.3	56.6
Precision	92.9	59.3	61.1
Recall	91.8	58.7	54.1
F1-score	92.3	58.5	56.6

Table 4.3 to Table 4.5 compare the TrGB and baseline cases for the AD diagnosis and for the three modality scenarios. As seen, while TrGB could improve the classification scores for all three situations, it has been most effective for the MRI modality case. The classification accuracy of the TrGB is 7%, 1.3%, and 2.8% higher than that of the baseline using MRI, PET, and multimodal data, respectively. When using the source model for target testing, the scores are slightly lower than those of the baseline. This situation can be compared with the CN/AD case of Table 4.2 where the source model was tested on the

source data. It can be concluded that the scores of the CN/AD classification using the source model decreased by almost 10% when the target domain was used as the test data. This shows the detrimental effect of data distribution shift between the source and target domains on the classification performance.

Table 4.3 CN/AD classification scores for the baseline, source-model, and TrGB scenarios for the MRI modality

	Baseline	Source-model	TrGB
Accuracy	72.0	70.3	78.7
Precision	70.9	68.7	77.8
Recall	71.1	70.4	78.3
F1-score	71.0	71.3	78.0

Table 4.4 CN/AD classification scores for the baselines, source-model, and TrGB scenarios for the PET modality

	Baseline	Source-model	TrGB
Accuracy	80.0	78.7	81.3
Precision	79.2	78.2	80.8
Recall	78.9	78.1	80.0
F1-score	79.0	79.3	80.3

Table 4.5 CN/AD classification scores for the baselines, source-model, and TrGB scenarios using the multimodal data

	Baseline	Source-model	TrGB
Accuracy	82.7	80.0	85.3
Precision	82.0	79.5	84.9
Recall	81.7	80.6	84.4
F1-score	81.8	79.6	84.6

Table 4.6 and Table 4.7 report the results for the two tasks of MCI and multiclass classification. Again, TrGB could improve the scores significantly. More specifically, the F1-score of the two tasks increased by 8.3% and 9.3%, respectively. It can also be seen that the source model has the worst performance when the MCI subjects are involved in the analysis. This outcome suggests that the MCI group of the two datasets has a more significant discrepancy in terms of marginal and conditional distributions. Figure 4.10 summarizes the results through F1-scores of the three methods (baseline, source-model, and TrGB) for the three classification tasks considered.

Table 4.6 CN/MCI classification scores for the baselines, source-model, and TrGB scenarios using the multimodal data

	Baseline	Source-model	TrGB
Accuracy	75.6	75.2	77.1
Precision	60.3	42.5	65.9
Recall	54.9	50.0	61.7
F1-score	54.6	46.0	62.9

Table 4.7 CN/MCI/AD classification scores for the baselines, source-model, and TrGB scenarios using the multimodal data

	Baseline	Source-model	TrGB
Accuracy	57.4	50.1	61.9
Precision	56.5	31.6	60.9
Recall	43.8	30.6	51.9
F1-score	45.2	30.7	54.5

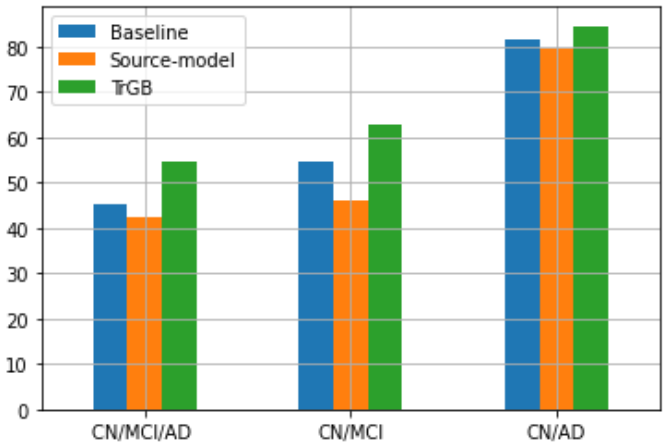


Figure 4.10 F1-score using the three methods, baselines, source-model, and TrGB for the three classification tasks, CN/MCI/AD, CN/MCI, and CN/AD, using multimodal data

Finally, an experiment for the early MCI vs. late MCI classification in the target domain was conducted by transferring the knowledge from the CN vs. AD classification of the source domain. The TrGB algorithm could enhance the performance of this task compared to the baseline. The results prove the effectiveness of the proposed transfer learning model for challenging classification tasks where the two classes are very similar, and the number of labeled instances is limited (Table 4.8).

Table 4.8 Early MCI vs. late MCI classification scores for the baselines and TrGB scenarios using the multimodal data

	Baseline	TrGB
Accuracy	57.4	60.4 3
Precision	46.9	53.2 6.3
Recall	47.9	52.8 4.9
F1-score	46.6	52.6 6

While the proposed transfer learning approach can boost the classification performance for a target dataset, there are some limitations associated with it. First, this approach increases the model complexity, which requires more training time and resources.

Also, since the model is based on GBM, the process of creating base learners is sequential, which is computationally intensive. For faster implementations and superior performance, alternative algorithms, including Extreme Gradient Boosting (XGB) and Light Gradient Boosting Machine (Light GBM), could be used. Another important circumstance in transfer learning techniques is negative transfer, where the transferred information from the source domain has a detrimental effect on the performance of the target model. In the proposed model, the defined weighting scheme aims to transfer as much information as possible from the source domain while avoiding negative transfer at the same time. Thus, selecting appropriate weights is essential to make a balance between these two conflicting requirements.

5. LONGITUDINAL PREDICTION OF AD USING RECURRENT NEURAL NETWORKS

5.1 Introduction

In addressing the barriers impeding AD research, scientists have proposed statistical and machine learning techniques for robust diagnosis. Until recently, most efforts were dedicated to modeling the disease at a single time point using cross-sectional datasets [79], [80]. However, these approaches could not provide enough information about the future status of patients. At later stages of AD, where the brain has already suffered from atrophy, treatment would be too late to be effective. Early diagnosis of the disease allows for early intervention and facilitates development of effective healthcare services. This initiates a new line of research aiming at enhancing the effect of treatment by predicting the onset of the disease before the occurrence of acute neurodegeneration. The objective of these studies is to leverage temporal information from longitudinal data to model the progression of AD. Multiple classification and regression models have been proposed to predict disease progression and level of disease severity. The feature space is either based on the information available at baseline or a concatenation of features from multiple previous time points [81]-[84]. The integration of features into a single observation window creates a high-dimensional input space which is not only difficult to deal with but also disregards temporal connections between consecutive time points [57], [85]. With the gradual nature of AD progression, these methods could not efficiently exploit the longitudinal information.

Recurrent Neural Networks (RNNs) were introduced in 1986 and recently gained more popularity. Using the inherent correlations of sequential data, RNNs proved their

potential in predicting AD-related biomarkers for a future time point. Although effective, these studies limit themselves to predicting only a single future interval. This work broadened the scope and application of the RNNs by predicting the progression of AD over multiple future time points simultaneously. Employing three records of data for each subject, the RNN surpassed other machine learning methods not only in estimating the categorical variable for a multiclass classification task, but also in assessing the numerical value of the AD biomarker. Furthermore, two variations of RNN, GRU and LSTM, are investigated for the challenging task of drawing the delineation boundary of subjects in a multiclass classification scenario and also for predicting the trajectories of cognitive scores for the next two years.

5.2 Methodology

5.2.1 Recurrent Neural Network (RNN)

Processing sequences of data, RNNs have the capability to effectively incorporate temporal dependencies in longitudinal data. Figure 5.1 illustrates an RNN with data sequences of k time steps. At each time point (t_i), besides the input features (X_{ti}), the internal state (memory) of the cell from the previous time step ($h_{t(i-1)}$) are fed to the cell. Thus, unlike feedforward neural networks, RNNs can identify patterns hidden in sequences of data. However, due to a lack of long-term memory in basic RNNs, each time point is mainly affected by previous intervals in close vicinity. Therefore, they are not capable of leveraging long-term relationships in historical data, and older information tends to fade away. This setback is known as “vanishing gradient” in which the network gradually forgets older traces.

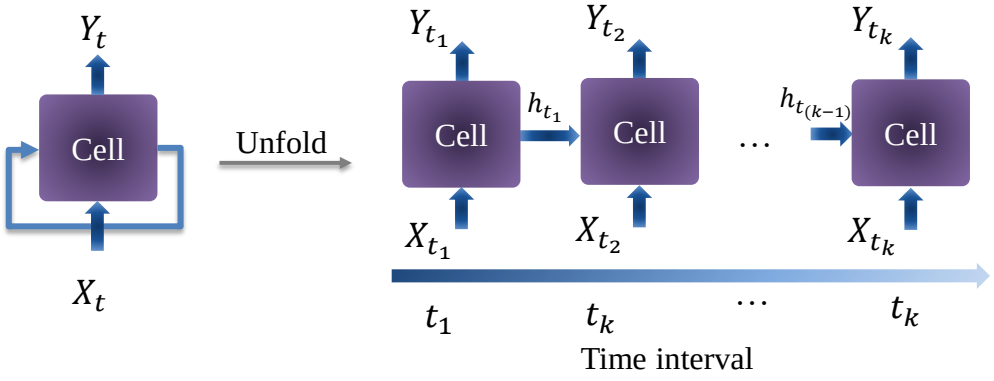


Figure 5.1 Recurrent Neural Network architecture.

To address this issue, GRU and LSTM-based RNN architectures with the capability of capturing long-term memories have been proposed [86], [87]. The structure of LSTM and GRU cells as the building blocks of an improved version of RNN is shown in Figure 5.2. In an LSTM cell, three gates denoted by sigmoid functions (σ) decide whether the previous cell state (C), the input (X), and the output (h) need to be passed to the next time step. This will make the memorizing capability of the cell more intelligent and durable. The following equations describe the operation principle of an LSTM cell.

$$
\begin{aligned}
f_{t_k} &= \sigma\left(W_f\left(X_{t_k}, h_{t_{k-1}}\right) + b_f\right) \\
i_{t_k} &= \sigma\left(W_i\left(X_{t_k}, h_{t_{k-1}}\right) + b_i\right) \\
\hat{\imath}_{t_k} &= tanh\left(W_i\left(X_{t_k}, h_{t_{k-1}}\right) + b_i\right) \\
C_{t_k} &= C_{t_{k-1}} * f_{t_k} + \hat{\imath}_{t_k} * i_{t_k} \\
o_{t_k} &= \sigma\left(W_o\left(X_{t_k}, h_{t_{k-1}}\right) + b_o\right) \\
h_{t_k} &= o_{t_k} * \tanh\left(C_{t_k}\right)
\end{aligned}
\tag{5-1}
$$

where t_k refers to the k^{th} time step; X_{t_k}, C_{t_k}, and h_{t_k} represent the input, state, and output of the cell at the k^{th} time step; and f_{t_k}, i_{t_k}, and o_{t_k} are the outputs of the forget, input, and output gates. Also, W and b are the weights of the neural networks.

In the gating mechanism of GRU, two gates known as reset and update gates determine the amount of the current input and output of the previous time step that needs to be preserved.

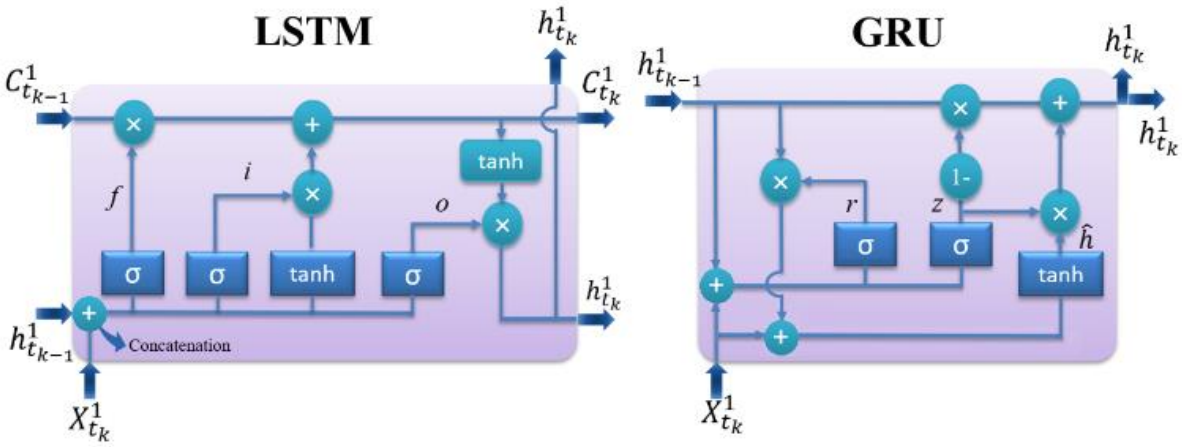


Figure 5.2 The structure of LSTM and GRU cells

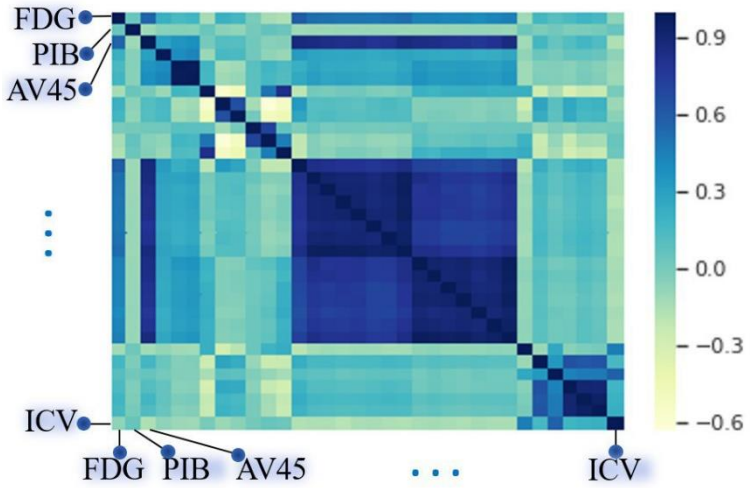


Figure 5.3 Heat-map of features used in this study

With the same notations of X_{t_k} and h_{t_k} as the input and output of the cell for the kth time step, the mathematical equations of a GRU cell are summarized as follows.

$$
\begin{aligned}
z_{t_k} &= \sigma\left(W_z\left(X_{t_k}, h_{t_{k-1}}\right) + b_z\right) \\
r_{t_k} &= \sigma\left(W_r\left(X_{t_k}, h_{t_{k-1}}\right) + b_r\right) \\
\hat{h}_{t_k} &= tanh\left(W_{\hat{h}}\left(X_{t_k}, r_{t_k} * h_{t_{k-1}}\right) + b_{\hat{h}}\right) \\
h_{t_k} &= \left(1 - z_{t_k}\right) * h_{t_{k-1}} + z_{t_k} * \hat{h}_{t_k} \\
h_{t_k} &= o_{t_k} * \tanh\left(C_{t_k}\right)
\end{aligned}
\tag{5-2}
$$

where z_{t_k} and r_{t_k} are the outputs of the update and reset gates.

5.2.2 Feature Selection

Referring to previous studies [59], which shed light on the possible overfitting of RNNs on the original feature space, feature analysis, and ranking has been performed on the data. Consequently, to address the highly correlated features, L1 feature selection was employed to extract the most important features. Using L1 regularization, 25 features with the highest variance in the feature space have been selected. The correlation matrix (heat map) of the features is illustrated in Figure 5.3.

5.2.3 Longitudinal AD Prediction using RNN

The proposed framework uses the memorization capability of the LSTM/GRU cell to capture historical dependencies from three records of subjects in order to predict the progression of AD at three next future time points. Therefore, a many-to-many RNN architecture with LSTM/GRU cells has been developed to carry out two tasks of longitudinal multiclass classification and regression.

The structure of the network for the LSTM case is demonstrated in Figure 5.4. In the developed network, the three inputs (X_{t_1}, $X_{t_2} and\ X_{t_3}$) represent the feature space associated with three-time points of M_0 (Baseline), M_6 (after 6 months), and M_{12} (after 12

months). The information is transferred from one time point to the next one using the cell state (C) and output (Y). The outputs Y_{t_i} are the Mini-Mental State Examination

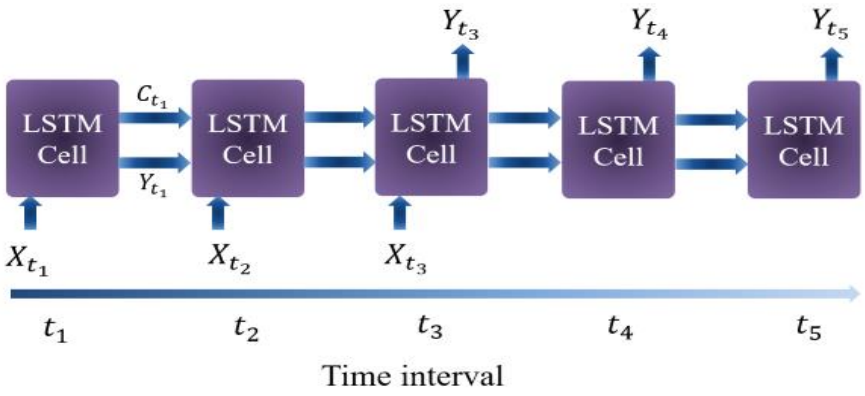


Figure 5.4 The RNN architecture used to predict the progression of AD using historical data

Table 5.1 Statistics of the Dataset Used in This Study

Category	Subjects (f/m)	Age	Education(y)	MMSE
AD	336 (150/186)	74.93±7.81	15.17 ±2.99	23.18 ± .06
MCI	864 (354/510)	73.03±7.60	15.91±2.85	27.59±1.81
CN	521 (268/253)	74.25±5.79	16.37±2.70	29.06±1.14

(MMSE) score for regression model or status of patients (CN, MCI, and AD) for classification model. The time steps t_4 and t_5 are associated with the future time points M_{24} (24 months after the baseline) and M_{36} (36 months after baseline). The next section discusses the material and experimental results.

5.3 Results and Discussion

5.3.1 Data

The data used in this study is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu/). 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 structural Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

Longitudinal medical records from 1458 subjects (341 CN, 255 EMCI, 529 LMCI, and 333 AD) have been incorporated in this dataset. During an 11-year study, each patient has been recalled for a follow-up visit every six months. These subjects have undergone several medical screening tests including MRI, PET, genetic tests, CSF tests, and cognitive impairment assessments. At each visit, an expert monitors the test results and updates the diagnosis for the participants. This categorical diagnosis (AD, MCI, NC) is used as the label for the multiclass classification experiment proposed in this study and the numerical Mini-Mental State Examination (MMSE) scores, an indicator of the AD cognitive impairment, with a range of 0-30 is adopted for the regression experiment. Characteristics of the dataset used in this study are summarized in Table 5.1.

5.3.2 Longitudinal Data Preprocessing

Initially, the data is preprocessed to alleviate any inconsistencies caused by utilizing different data modalities and various protocols. Subjects who have participated at all five

consecutive intervals including baseline, six months after the first visit (M_{06}), twelve months after the first visit (M_{12}), twenty-four months after the first visit (M_{24}), and thirty-six months after the first visit (M_{36}) have been considered. In the initial step of the experiments, data cleaning [88], [89], mean centering, data

Table 5.2 Summary Of Multimodal Features Utilized In This Study

Source	Features
Cognitive tests	Everyday Cognition (ECog) questionnaire measurements, FAQ, MOCA, RAVLT, CDRSB
MRI	Ventricular volume, Hippocampus volume, Whole Brain volume, Entorhinal Cortical thickness, Fusiform, Middle temporal gyrus, ICV
PET	FDG, PIB amyloid, AV45 amyloid
Genetic	APOE4
Demographic	Age, Gender, Education
CSF	Amyloid Beta, Phosphorylated Tau, Total Tau

Table 5.3 Regression Results

Algorithm	M12		M24		M36		Total
	RMSE	Corr	RMSE	Corr	RMSE	Corr	MSE
Ridge	2.07	0.58	2.66	0.62	2.99	0.63	6.82
SVR	2.14	0.59	2.86	0.61	3.17	0.58	7.68
LSTM	1.97	**0.63**	2.33	0.69	2.54	**0.72**	5.26
GRU	1.97	**0.63**	2.33	0.69	2.54	**0.72**	5.24
Ridge + FS*	2.02	0.62	2.67	0.65	2.93	0.65	6.65
SVR + FS*	2.16	0.60	2.76	0.65	3.26	0.62	7.70
LSTM + FS*	1.85	**0.63**	2.25	0.70	2.48	0.70	4.98
GRU + FS*	**1.82**	**0.63**	**2.21**	**0.71**	**2.44**	0.70	**4.77**

*Feature selection

normalization, missing feature handling, and univariate feature analysis has been performed to discard uninformative features. Furthermore, subjects whose medical diagnosis are not reported are removed from further analysis.

5.3.3 Simulation and Results

This study evaluates the performance of two RNN variations, LSTM and GRU, on the ADNI cohort for the two tasks of classification and regression. The experiment proceeds with the selection of historical records from subjects at three intervals (baseline, M_{06}, and M_{12}) to predict the status of the subjects in three future time points of M_{12}, M_{24} and M_{36}. Estimating the MMSE scores of subjects is pursued as a regression problem and predicting the diagnosis labels is defined as multiclass classification problem. The data has been split randomly to a 75% training set, a 10% validation set, and a 15% testing set. Grid search has been utilized to select the best hyperparameters for regression and classification networks separately. In order to feed the longitudinal feature space into the RNNs, the data has been framed in the tensor form of [samples, time steps, features], which in this case is 3-time steps of the 532 samples with 34 features involving MRI, PET, Cerebrospinal fluid (CSF) and cognitive test scores as provided in Table 5.2.

The performance of LSTM and GRU, implemented using the Keras deep learning library, are compared with state-of-the-art methods. It is worth noting that conventional methods cannot incorporate historical records of subjects to enhance prediction accuracy. This limitation has been compensated by concatenating all three-historical feature sets. Competing methods are then trained on this new feature space to find an individual direct

map between the feature space from past intervals with the corresponding future time points.

As regression, RMSE and R-Correlation factor are used as evaluation metrics to compare Ridge and SVR from Scikit-learn library with LSTM and GRU, and the Results are reported in Table 5.3.

Table 5.4 Classification Results

Method	M12			
	ACC	PRE	REC	F1
SVM	0.66±0.04	0.44±0.05	0.66±0.05	0.52±0.04
LSTM	0.84±0.10	0.86±0.06	0.84±0.10	0.81±0.16
GRU	0.61±0.09	0.95±0.00	0.60±0.09	0.74±0.07
LSTM + FS	0.88±0.03	0.89±0.02	0.90±0.02	0.89±0.02
GRU + FS	0.68±0.09	0.95±0.00	0.68±0.09	0.79±0.07
Method	M24			
	ACC	ACC	ACC	ACC
SVM	0.61±0.04	0.61±0.04	0.61±0.04	0.61±0.04
LSTM	0.82±0.12	0.82±0.12	0.82±0.12	0.82±0.12
GRU	0.37±0.06	0.37±0.06	0.37±0.06	0.37±0.06
LSTM + FS	0.87±0.01	0.87±0.01	0.87±0.01	0.87±0.01
GRU + FS	0.28±0.11	0.28±0.11	0.28±0.11	0.28±0.11
Method	M36			
	ACC	PRE	REC	F1
SVM	0.61±0.03	0.38±0.04	0.61±0.03	0.48±0.04
LSTM	0.80±0.09	0.84±0.06	0.80±0.09	0.78±0.15
GRU	0.61±0.04	0.98±0.00	0.61±0.04	0.75±0.03
LSTM + FS	0.88±0.02	0.87±0.03	0.88±0.02	0.87±0.03
GRU + FS	0.51±0.08	0.98±0.00	0.51±0.08	0.67±0.04

Similarly, the classification problem is defined as the diagnosis of subjects at three future time points based on three previous intervals. For the classification task, SVM from Scikit-learn library is selected as the competitive alternative to evaluate performance of the LSTM and GRU. F-score, precision, recall, accuracy has been utilized as the classification metrics and the results are summarized in Table 5.4.

From Table 5.3 and Table 5.4, it can be observed that the LSTM and GRU on the original feature space demonstrate lower performance in comparison to the competitive methods in some cases. Incorporating L1 has led to a noticeable improvement in the prediction accuracy, which could be associated with the overfitting of networks. Since RNNs have a high number of variables and weights, they require a larger number of samples for training. The approach investigated in this work employs the L1 feature selection to overcome the limited number of samples for training an effective network, which can predict the future status of AD subjects using their historical measurements.

6. CONCLUSION

In this dissertation, various machine learning frameworks have been explored for multimodal feature engineering and multiclass classification of mild cognitive impairment (MCI) and Alzheimer's disease (AD). Moreover, to overcome data shortage in smaller datasets, a transfer learning technique has been developed to combine multiple data domains with different data distributions.

A feature selection approach based on information theory-based metrics has been proposed for multimodal neuroimaging biomarkers, including amyloid and tau PET SUVR values and cortical thickness. A trade-off was made between the relevance and redundancy of the attributes using the mutual information metric. The advantage of the proposed approach is to incorporate the effect of feature complementarity and redundancy to maximize the total amount of relevant information in the feature set. This will also reduce the computational burden while preserving the accuracy in the classification and prediction results. It is important to note that the redundancy part should not be overweighted since highly relevant features can also be partially redundant. This situation was depicted in Figure 2.4 for larger values of the coefficient α, where feature relevance is sacrificed for even a minor redundancy.

In the classification part, tau PET modality produced more accurate results than amyloid PET and MRI modalities, specifically in the most challenging CN/MCI classification (early stage). On the other hand, multimodal scenarios have achieved the highest F1-scores in most cases, especially in the critical early stages of the disease. Feature selection was most effective in the SVC case, making SVC achieve higher scores compared

to RF and XGB in many cases. This was expected as RF and XGB have internal feature selection, with less room for improvement. In retrospect, these findings suggest that the classification of high-dimensional multimodal datasets would be most accurate when feature selection is carried out most effectively, with the relevance of each feature quantified through a ranking score metric as proposed in this dissertation. When such measures are taken, reducing the dimensionality of the feature space can be accomplished while still maintaining high accuracy in the classification results. More specifically, Fig. 9 (d) showed that the F1-score of the multimodal case with feature selection is up to 5% higher than in other scenarios.

One of the major challenges in the AD diagnosis is the heterogeneity of the disease related to the AD subtypes (hippocampal-sparing, limbic-predominant, typical AD). It is shown that the AD risk factors as well as the protective factors have a meaningful variance among the AD subtypes [90]. As seen in the result section, the inclusion of these covariates into the model variables could improve the classification scores. This can be explained through the characteristics of different subtypes and the variation of risk factors among them. Typical AD subtype cases experience more severe pathology compared to other subtypes, while limbic-predominant cases have more typical biomarkers than hippocampal-sparing subjects. Since typical AD is more prevalent than other subtypes, if the classification model only relies on biomarkers, it might be biased toward this group and hence could yield false-negative results for other AD subtypes as they have less severe biomarkers and are also less prevalent. Therefore, these other categories of subjects with minimal atrophy and non-typical biomarkers might be misclassified as CN and MCI classes. At this stage, these risk and protective factors can complement the biomarkers and

help to correctly classify these subtypes as the AD group and thus alleviate the complexity of the heterogeneity issue. Concerning the risk factors, subjects with typical and limbic-predominant AD tend to be older than those with hippocampal-sparing AD. On the other hand, the hippocampal-sparing category includes fewer APOE4 carriers and highly educated individuals compared to other groups. In terms of gender, females are more frequent in the limbic-predominant group.

Since the biomarkers might not be accessible in many situations, clinical diagnosis is made solely based on symptoms as ascertained through cognitive tests. The AT(N) biomarker framework establishes a biomarker-based definition of AD and emphasizes the independence of the biological and clinical definitions of AD, yet it tries to clarify the interaction between the two. This can be valuable for in-depth research purposes as well as personalized medicine. The AT(N) framework shows that the cognitive stage cannot be entirely determined through the AT(N) biomarkers since any particular biomarker profile can belong to any cognitive stage. The fact that a wide range of biomarker profiles can define a specific cognitive stage is due to the heterogeneity of the disease, which can be explained by the subtypes of AD (hippocampal-sparing, limbic-predominant, typical AD). Different subtypes have similar amyloid loads; however, tau and neurodegeneration pathology and concomitant non-AD pathologies vary across subtypes. Also, other contributing factors to differentiate between AD subtypes include risk factors (age, gender, education, and APOE) and protective factors (cognitive reserve, brain resilience, and brain resistance). Assimilation of these factors in the context of the AT(N) system can be a step forward for a more in-depth analysis of the computer-aided diagnosis of AD and for augmenting the research prospects for more effective personalized medicine.

One of the limiting factors faced in our analysis was the considerable amount of missing data, specifically for the tau PET modality. This issue is more critical when we are interested in subjects with all modalities being available, which is a requirement for having a fair comparison between single modality scenarios. Also, the study could be more valuable if longitudinal data were available so that the effect of biomarker change through time could be considered. Longitudinal tau PET data is very limited in the ADNI dataset since tau PET is a relatively new technology, and its longitudinal data collection and processing are still in progress. Also, the missing data issue is even more severe for the longitudinal data.

Moreover, in the data collection process, a time difference may exist between capturing the MRI and PET scans for some participants. This time lag between modalities is inevitable in many follow up situations in medical practice. While small time-lags might be neglected in some studies, more significant delays can be included in the analysis with appropriate considerations. In our study, we have not integrated this variable in our analysis due to the lack of such information for some of the participants, which would result in additional missing values for the dataset. In this study, we conducted a cross-sectional study and handled the missing values by mean-value imputation and by making use of models that are more robust to missing values. Moreover, using the AT(N) analysis, the intra-class biomarker variance was studied so that the contribution of biomarker shortage on the classification performance was determined.

To transfer knowledge from an auxiliary dataset to a target dataset with insufficient training data for a classification task, a novel transfer learning framework based on the gradient boosting machine (TrGB) was proposed. Like the GBM algorithm, the TrGB base

learners focus on the residual of their previous learners. In addition, the residuals are used to assign a weight to the source and target instances such that misleading instances of the source domain have a more negligible effect on the learning process. The experimental results show that the TrGB achieved superior performance for MCI and AD classification compared to the baseline case, where a model is trained only using the labeled instances of the target domain. It was shown that the MCI group has more data distribution shift between the two datasets than the AD group. Transfer learning was also beneficial for the early MCI vs. late MCI classification in the target dataset using the knowledge in the CN vs. AD classification model of the source domain.

For tracking the progression of the AD at multiple future intervals and gauging the merits and gradual effects of any potential treatment plan in longitudinal AD studies, an approach based on Recurrent Neural Networks (RNN) was presented. Three historical time points from subjects in three categories of CN, MCI, and AD were selected to form a feature space. Then, the model is trained on 75% of the data to predict three future MMSE scores and diagnosis labels of the subjects with two different variations of RNN (LSTM and GRU). It was observed that by employing L1 feature extraction prior to application of the RNNs lead to a higher performance in both regression and classification models in comparison to other state-of-the-art algorithms, as supported by the results provided in Table 5.3 and Table 5.4.

REFERENCES

[1] 2019 Alzheimer's Disease Facts and Figures. https://www.alz.org/media/Documents/alzheimers-facts-and-figures.pdf

[2] S. Tabarestani, M. Aghili, M. Eslami, M. Cabrerizo, A. Barreto, N. Rishe, R. Curiel, D. Loewenstein, R. Duara and M. Adjouadi, "A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study", NeuroImage, vol. 206, p. 116317, 2020.

[3] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. Fulham and ADNI, "Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease", IEEE Transactions on Biomedical Engineering, vol. 62, no. 4, pp. 1132-1140, 2015.

[4] M. Ewers, R. Sperling, W. Klunk, M. Weiner and H. Hampel, "Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia", Trends in Neurosciences, vol. 34, no. 8, pp. 430-442, 2011.

[5] B. Dubois, H. Hampel, H. Feldman, P. Scheltens, P. Aisen, S. Andrieu, H. Bakardjian, H. Benali, L. Bertram, K. Blennow, K. Broich, E. Cavedo, S. Crutch, J. Dartigues, C. Duyckaerts, S. Epelbaum, G. Frisoni, S. Gauthier, R. Genthon, A. Gouw, M. Habert, D. Holtzman, M. Kivipelto, S. Lista, J. Molinuevo, S. OBryant, G. Rabinovici, C. Rowe, S. Salloway, L. Schneider, R. Sperling, M. Teichmann, M. Carrillo, J. Cummings and C. Jack, "Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria", Alzheimer's&Dementia, vol.12, no.3, pp.292-323, 2016.

[6] D. Loewenstein, R. Curiel, S. DeKosky, R. Bauer, M. Rosselli, S. Guinjoan, M. Adjouadi, A. Peñate, W. Barker, S. Goenaga, T. Golde, M. Greig-Custo, K. Hanson, C. Li, G. Lizarraga, M. Marsiske and R. Duara, "Utilizing semantic intrusions to identify amyloid positivity in mild cognitive impairment", Neurology, vol. 91, no. 10, pp. e976-e984, 2018.

[7] M. Albert, S. DeKosky, D. Dickson, B. Dubois, H. Feldman, N. Fox, A. Gamst, D. Holtzman, W. Jagust, R. Petersen, P. Snyder, M. Carrillo, B. Thies and C. Phelps, "The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease", Alzheimer's & Dementia, vol. 7, no. 3, pp. 270-279, 2011.

[8] C. MISRA, Y. FAN and C. DAVATZIKOS, "Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI", NeuroImage, vol. 44, no. 4, pp. 1415-1422, 2009.

[9] D. Selkoe, "Early network dysfunction in Alzheimer's disease", Science, vol. 365, no. 6453, pp. 540-541, 2019.

[10] G. Ridgway, M. Lehmann, J. Barnes, J. Rohrer, J. Warren, S. Crutch and N. Fox, "Early-onset Alzheimer disease clinical variants: Multivariate analyses of cortical thickness", Neurology, vol. 79, no. 1, pp. 80-84, 2012.

[11] M. Bergeron, S. Landset, X. Zhou, T. Ding, T. Khoshgoftaar, F. Zhao, B. Du, X. Chen, X. Wang, L. Zhong, X. Liu and J. Ashford, "Utility of MemTrax and Machine Learning Modeling in Classification of Mild Cognitive Impairment", Journal of Alzheimer's Disease, vol. 77, no. 4, pp. 1545-1558, 2020.

[12] A. Ezzati, A. Zammit, D. Harvey, C. Habeck, C. Hall and R. Lipton, "Optimizing Machine Learning Methods to Improve Predictive Models of Alzheimer's Disease", Journal of Alzheimer's Disease, vol. 71, no. 3, pp. 1027-1036, 2019.

[13] S. Gill, P. Mouches, S. Hu, D. Rajashekar, F. MacMaster, E. Smith, N. Forkert and Z. Ismail, "Using Machine Learning to Predict Dementia from Neuropsychiatric Symptom and Neuroimaging Data", Journal of Alzheimer's Disease, vol. 75, no. 1, pp. 277-288, 2020.

[14] F. Vecchio, F. Miraglia, F. Alù, M. Menna, E. Judica, M. Cotelli and P. Rossini, "Classification of Alzheimer's Disease with Respect to Physiological Aging with Innovative EEG Biomarkers in a Machine Learning Implementation", Journal of Alzheimer's Disease, vol. 75, no. 4, pp. 1253-1261, 2020.

[15] V. Villemagne, M. Fodero-Tavoletti, C. Masters and C. Rowe, "Tau imaging: early progress and future directions", The Lancet Neurology, vol. 14, no. 1, pp. 114-124, 2015.

[16] K. Johnson, A. Schultz, R. Betensky, J. Becker, J. Sepulcre, D. Rentz, E. Mormino, J. Chhatwal, R. Amariglio, K. Papp, G. Marshall, M. Albers, S. Mauro, L. Pepin, J. Alverio, K. Judge, M. Philiossaint, T. Shoup, D. Yokell, B. Dickerson, T. Gomez-Isla, B. Hyman, N. Vasdev and R. Sperling, "Tau positron emission tomographic imaging in aging and early Alzheimer disease", Annals of Neurology, vol. 79, no. 1, pp. 110-119, 2015.

[17] L. Ittner and J. Götz, "Amyloid-β and tau — a toxic pas de deux in Alzheimer's disease", Nature Reviews Neuroscience, vol. 12, no. 2, pp. 67-72, 2010.

[18] A. Leuzy, K. Chiotis, L. Lemoine, P. Gillberg, O. Almkvist, E. Rodriguez-Vieitez and A. Nordberg, "Tau PET imaging in neurodegenerative tauopathies—still a challenge", Molecular Psychiatry, vol. 24, no. 8, pp. 1112-1134, 2019.

[19] F. de Wolf, M. Ghanbari, S. Licher, K. McRae-McKee, L. Gras, G. Weverling, P. Wermeling, S. Sedaghat, M. Ikram, R. Waziry, W. Koudstaal, J. Klap, S. Kostense, A. Hofman, R. Anderson, J. Goudsmit and M. Ikram, "Plasma tau, neurofilament light chain and amyloid-β levels and risk of dementia; a population-based cohort study", Brain, vol. 143, no. 4, pp. 1220-1232, 2020.

[20] M. Schöll and A. Maass, "Does early cognitive decline require the presence of both tau and amyloid-β?", Brain, vol. 143, no. 1, pp. 10-13, 2019.

[21] R. Ossenkoppele, R. Smith, T. Ohlsson, O. Strandberg, N. Mattsson, P. Insel, S. Palmqvist and O. Hansson, "Associations between tau, Aβ, and cortical thickness with cognition in Alzheimer disease", Neurology, vol. 92, no. 6, pp. e601-e612, 2019.

[22] B. Jedynak, A. Lang, B. Liu, E. Katz, Y. Zhang, B. Wyman, D. Raunig, C. Jedynak, B. Caffo and J. Prince, "A computational neurodegenerative disease progression score: Method and results with the Alzheimer's disease neuroimaging initiative cohort", NeuroImage, vol. 63, no. 3, pp. 1478-1486, 2012.

[23] C. Jack, D. Knopman, W. Jagust, L. Shaw, P. Aisen, M. Weiner, R. Petersen and J. Trojanowski, "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade", The Lancet Neurology, vol. 9, no. 1, pp. 119-128, 2010.

[24] M. Shojaie, S. Tabarestani, M. Cabrerizo, S. DeKosky, D. Vaillancourt, D. Loewenstein, R. Duara and M. Adjouadi, "PET Imaging of Tau Pathology and Amyloid-β, and MRI for Alzheimer's Disease Feature Fusion and Multimodal Classification", Journal of Alzheimer's Disease, vol. 84, no. 4, pp. 1497-1514, 2021.

[25] S. Lee, Z. Xu, T. Li and Y. Yang, "A novel bagging C4.5 algorithm based on wrapper feature selection for supporting wise clinical decision making", Journal of Biomedical Informatics, vol. 78, pp. 144-155, 2018.

[26] F. Zhuang et al., "A Comprehensive Survey on Transfer Learning", Proceedings of the IEEE, vol. 109, no. 1, pp. 43-76, 2021.

[27] Z. Wang, Z. Dai, B. Poczos and J. Carbonell, "Characterizing and Avoiding Negative Transfer", 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[28] D. Zhang, Y. Wang, L. Zhou, H. Yuan and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment", NeuroImage, vol. 55, no. 3, pp. 856-867, 2011.

[29] Y. Shi, H. Suk, Y. Gao, S. Lee and D. Shen, "Leveraging Coupled Interaction for Multimodal Alzheimer's Disease Diagnosis", IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 1, pp. 186-200, 2020.

[30] S. Tabarestani, M. Aghili, M. Shojaie, C. Freytes and M. Adjouadi, "Profile-Specific Regression Model for Progression Prediction of Alzheimer's Disease Using Longitudinal Data", 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018.

[31] S. Tabarestani, M. Aghili, M. Shojaie, C. Freytes, M. Cabrerizo, A. Barreto, N. Rishe, R. Curiel, D. Loewenstein, R. Duara and M. Adjouadi, "Longitudinal Prediction Modeling of Alzheimer Disease using Recurrent Neural Networks", 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2019.

[32] L. An, E. Adeli, M. Liu, J. Zhang, S. Lee and D. Shen, "A Hierarchical Feature and Sample Selection Framework and Its Application for Alzheimer's Disease Diagnosis", Scientific Reports, vol. 7, no. 1, 2017.

[33] J. Shi, X. Zheng, Y. Li, Q. Zhang and S. Ying, "Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease", IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 1, pp. 173-183, 2018.

[34] J. Vergara and P. Estévez, "A review of feature selection methods based on mutual information", Neural Computing and Applications, vol. 24, no. 1, pp. 175-186, 2013.

[35] Y. Saeys, I. Inza and P. Larranaga, "A review of feature selection techniques in bioinformatics", Bioinformatics, vol. 23, no. 19, pp. 2507-2517, 2007.

[36] X. Hao, Y. Bao, Y. Guo, M. Yu, D. Zhang, S. Risacher, A. Saykin, X. Yao and L. Shen, "Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease", Medical Image Analysis, vol. 60, p. 101625, 2020.

[37] I. Garali, M. Adel, S. Bourennane and E. Guedj, "Histogram-Based Features Selection and Volume of Interest Ranking for Brain PET Image Classification", IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, pp. 1-12, 2018.

[38] X. Zhu, H. Suk, S. Lee and D. Shen, "Subspace Regularized Sparse Multitask Learning for Multiclass Neurodegenerative Disease Identification", IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 607-618, 2016.

[39] M. Liu, D. Zhang and D. Shen, "Relationship Induced Multi-Template Learning for Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment", IEEE Transactions on Medical Imaging, vol. 35, no. 6, pp. 1463-1474, 2016.

[40] L. Xu, Z. Yao, J. Li, C. Lv, H. Zhang and B. Hu, "Sparse Feature Learning With Label Information for Alzheimer's Disease Classification Based on Magnetic Resonance Imaging", IEEE Access, vol. 7, pp. 26157-26167, 2019.

[41] P. Jiang, X. Wang, Q. Li, L. Jin and S. Li, "Correlation-Aware Sparse and Low-Rank Constrained Multi-Task Learning for Longitudinal Analysis of Alzheimer's Disease", IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 4, pp. 1450-1456, 2019.

[42] C. Jack, D. Bennett, K. Blennow, M. Carrillo, B. Dunn, S. Haeberlein, D. Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liu, J. Molinuevo, T. Montine, C. Phelps, K. Rankin, C. Rowe, P. Scheltens, E. Siemers, H. Snyder, R. Sperling, C. Elliott, E. Masliah, L. Ryan and N. Silverberg, "NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease", Alzheimer's & Dementia, vol. 14, no. 4, pp. 535-562, 2018.

[43] S. Pan and Q. Yang, "A Survey on Transfer Learning", IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2010.

[44] W. Dai, Q. Yang, G. Xue and Y. Yu, "Boosting for transfer learning", Proceedings of the 24th international conference on Machine learning - ICML '07, 2007.

[45] Y. Yao and G. Doretto, "Boosting for transfer learning with multiple sources", 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010.

[46] Q. Zhang, H. Li, Y. Zhang and M. Li, "Instance Transfer Learning with Multisource Dynamic TrAdaBoost", The Scientific World Journal, vol. 2014, pp. 1-8, 2014.

[47] Y. Yang, X. Li, P. Wang, Y. Xia and Q. Ye, "Multi-Source Transfer Learning via Ensemble Approach for Initial Diagnosis of Alzheimer's Disease", IEEE Journal of Translational Engineering in Health and Medicine, vol. 8, pp. 1-10, 2020.

[48] N. Khan, N. Abraham and M. Hon, "Transfer Learning With Intelligent Training Data Selection for Prediction of Alzheimer's Disease", IEEE Access, vol. 7, pp. 72726-72735, 2019.

[49] A. Mehmood, S. Yang, Z. Feng, M. Wang, A. S. Ahmad, R. Khan, M. Maqsood, and M. Yaqub, "A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images", Neuroscience, vol. 460, pp. 43-52, 2021.

[50] H. Shin, H. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. Summers, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning", IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285-1298, 2016.

[51] Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, and Summers RM, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," IEEE Trans Med Imaging 35(5):1285–1298, 2016.

[52] M. Maqsood, F. Nazir, U. Khan, F. Aadil, H. Jamal, I. Mehmood, and O. Song, "Transfer Learning Assisted Classification and Detection of Alzheimer's Disease Stages Using 3D MRI Scans", Sensors, vol. 19, no. 11, p. 2645, 2019.

[53] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, "Classification of alzheimer disease on imaging modalities with deep cnns using cross-modal transfer learning," 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), 2018.

[54] Lixin Duan, Dong Xu, and I. W. Tsang, "Domain adaptation from multiple sources: A domain-dependent regularization approach," IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 3, pp. 504–518, 2012.

[55] R. Cui, M. Liu, and G. Li, "Longitudinal analysis for Alzheimer's disease diagnosis using RNN," Proc. - Int. Symp. Biomed. Imaging, vol. 2018–April, no. Isbi, pp. 1398–1401, 2018.

[56] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, “Using recurrent neural network models for early detection of heart failure onset,” J. Am. Med. Informatics Assoc., vol. 24, no. 2, pp. 361–370, 2017.

[57] M. Nguyen, N. Sun, D. C. Alexander, J. Feng, and B. T. Thomas Yeo, “Modeling Alzheimer’s disease progression using deep recurrent neural networks,” in 2018 International Workshop on Pattern Recognition in Neuroimaging, PRNI 2018, 2018.

[58] T. Wang, R. G. Qiu, and M. Yu, “Predictive Modeling of the Progression of Alzheimer ’ s Disease with Recurrent Neural Networks,” Sci. Rep., pp. 1–12, 2018.

[59] M. Aghili, S. Tabarestani, M. Adjouadi and E. Adeli, "Predictive Modeling of Longitudinal Data for Alzheimer’s Disease Diagnosis Using RNNs", PRedictive Intelligence in MEdicine, pp. 112-119, 2018.

[60] X. Zhu, H. Suk, L. Wang, S. Lee and D. Shen, "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis", Medical Image Analysis, vol. 38, pp. 205-214, 2017.

[61] J. Tohka, E. Moradi and H. Huttunen, "Comparison of Feature Selection Techniques in Machine Learning for Anatomical Brain MRI in Dementia", Neuroinformatics, vol. 14, no. 3, pp. 279-296, 2016.

[62] X. Zhu, H. Suk, S. Lee and D. Shen, "Canonical feature selection for joint regression and multi-class identification in Alzheimer’s disease diagnosis", Brain Imaging and Behavior, vol. 10, no. 3, pp. 818-828, 2015.

[63] S. Mishra, B. Gordon, Y. Su, J. Christensen, K. Friedrichsen, K. Jackson, R. Hornbeck, D. Balota, N. Cairns, J. Morris, B. Ances and T. Benzinger, "AV-1451 PET imaging of tau pathology in preclinical Alzheimer disease: Defining a summary measure", NeuroImage, vol. 161, pp. 171-178, 2017.

[64] L. Jorge, R. Martins, N. Canário, C. Xavier, A. Abrunhosa, I. Santana and M. Castelo-Branco, "Investigating the Spatial Associations Between Amyloid-β Deposition, Grey Matter Volume, and Neuroinflammation in Alzheimer's Disease", Journal of Alzheimer's Disease, vol. 80, no. 1, pp. 113-132, 2021.

[65] W. Wang, J. Yu, Y. Liu, R. Yin, H. Wang, J. Wang, L. Tan, J. Radua and L. Tan, "Voxel-based meta-analysis of grey matter changes in Alzheimer's disease", Translational Neurodegeneration, vol. 4, no. 1, 2015.

[66] C. Jack, H. Wiste, C. Schwarz, V. Lowe, M. Senjem, P. Vemuri, S. Weigand, T. Therneau, D. Knopman, J. Gunter, D. Jones, J. Graff-Radford, K. Kantarci, R. Roberts, M. Mielke, M. Machulda and R. Petersen, "Longitudinal tau PET in ageing and Alzheimer's disease", Brain, vol. 141, no. 5, pp. 1517-1528, 2018.

[67] A. Maass, S. Landau, S. Baker, A. Horng, S. Lockhart, R. La Joie, G. Rabinovici and W. Jagust, "Comparison of multiple tau-PET measures as biomarkers in aging and Alzheimer's disease", NeuroImage, vol. 157, pp. 448-463, 2017.

[68] V. Lowe, H. Wiste, M. Senjem, S. Weigand, T. Therneau, B. Boeve, K. Josephs, P. Fang, M. Pandey, M. Murray, K. Kantarci, D. Jones, P. Vemuri, J. Graff-Radford, C. Schwarz, M. Machulda, M. Mielke, R. Roberts, D. Knopman, R. Petersen and C. Jack, "Widespread brain tau and its association with ageing, Braak stage and Alzheimer's dementia", Brain, vol. 141, no. 1, pp. 271-287, 2017.

[69] Y. Liu, W. Yan, C. Tan, J. Li, W. Xu, X. Cao, L. Tan and J. Yu, "Common Variant in TREM1 Influencing Brain Amyloid Deposition in Mild Cognitive Impairment and Alzheimer's Disease", Neurotoxicity Research, vol. 37, no. 3, pp. 661-668, 2019.

[70] G. Chetelat, V. Villemagne, P. Bourgeat, K. Pike, G. Jones, D. Ames, K. Ellis, C. Szoeke, R. Martins, G. O'Keefe, O. Salvado, C. Masters and C. Rowe, "Relationship between atrophy and B-amyloid deposition in Alzheimer's disease", Annals of Neurology, p. NA-NA, 2009.

[71] E. Crocco, D. Loewenstein, R. Curiel, N. Alperin, S. Czaja, P. Harvey, X. Sun, J. Lenchus, A. Raffo, A. Peñate, J. Melo, L. Sang, R. Valdivia and K. Cardenas, "A novel cognitive assessment paradigm to detect Pre-mild cognitive impairment (PreMCI) and the relationship to biological markers of Alzheimer's disease", Journal of Psychiatric Research, vol. 96, pp. 33-38, 2018.

[72] G. Schwindt and S. Black, "Functional imaging studies of episodic memory in Alzheimer's disease: a quantitative meta-analysis", NeuroImage, vol. 45, no. 1, pp. 181-190, 2009.

[73] S. Landau, M. Mintun, A. Joshi, R. Koeppe, R. Petersen, P. Aisen, M. Weiner and W. Jagust, "Amyloid deposition, hypometabolism, and longitudinal cognitive decline", Annals of Neurology, vol. 72, no. 4, pp. 578-586, 2012.

[74] C. Jack, H. Wiste, S. Weigand, T. Therneau, V. Lowe, D. Knopman, J. Gunter, M. Senjem, D. Jones, K. Kantarci, M. Machulda, M. Mielke, R. Roberts, P. Vemuri, D. Reyes and R. Petersen, "Defining imaging biomarker cut points for brain aging and Alzheimer's disease", Alzheimer's & Dementia, vol. 13, no. 3, pp. 205-216, 2016.

[75] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Schölkopf, "Correcting sample selection bias by unlabeled data," in Proc. 20th Annu. Conf. Neural Inf. Process. Syst., Vancouver, BC, Canada, Dec. 2006, pp. 601–608, 2006.

[76] M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. von Bünau, and M. Kawanabe, “Direct importance estimation for covariate shift adaptation,” Annals of the Institute of Statistical Mathematics, vol. 60, no. 4, pp. 699–746, 2008.

[77] J. Juntu, J. Sijbers, D. Van Dyck, and J. Gielen, “Bias field correction for MRI images,” in Proc. Comput. Recognit. Syst. Springer, pp. 543–551, 2005.

[78] J. H. Friedman, “Greedy function approximation: A gradient boosting machine.,” The Annals of Statistics, vol. 29, no. 5, 2001.

[79] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung, and S. C. Johnson, “Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset,” Neuroimage, vol. 48, no. 1, pp. 138–149, 2009.

[80] Y. Gao, C. Wee, M. Kim, P. Giannakopoulos, M. Montandon, S. Haller and D. Shen, "MCI Identification by Joint Learning on Multiple MRI Data", Lecture Notes in Computer Science, pp. 78-85, 2015.

[81] J. Zhou, J. Liu, V. A. Narayan, and J. Ye, “Modeling disease progression via multi-task learning,” Neuroimage, vol. 78, pp. 233–248, 2013.

[82] L. Huang, Y. Jin, Y. Gao, K.-H. Thung, and D. Shen, “Longitudinal clinical score prediction in Alzheimer’s disease with soft-split sparse regression based random forest,” Neurobiol. Aging, vol. 46, pp. 180–191, 2016.

[83] G. Gavidia-Bovadilla, S. Kanaan-Izquierdo, M. Mataró-Serrat and A. Perera-Lluna, "Early Prediction of Alzheimer’s Disease Using Null Longitudinal Model-Based Classifiers", PLOS ONE, vol. 12, no. 1, p. e0168011, 2017.

[84] S. Tabarestani, M. Aghili, M. Shojaie, C. Freytes and M. Adjouadi, "Profile-Specific Regression Model for Progression Prediction of Alzheimer's Disease Using Longitudinal Data," *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1353-1357, 2018.

[85] S. Tsao, N. Gajawelli, J. Zhou, J. Shi, J. Ye, , Y. Wang & N. Leporé, “Feature selective temporal prediction of Alzheimer’s disease progression using hippocampus surface morphometry,” Brain Behav., vol. 7, no. 7, pp. 1–11, 2017.

[86] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Comput., 1997.

[87] K. Cho, B. van Merrienboer, Ç. Gülçehre, F. Bougares, H. Schwenk, and Y. Bengio, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” CoRR, vol. abs/1406.1078, 2014.

[88] M. Mafi, S. Tabarestani, M. Cabrerizo, A. Barreto and M. Adjouadi, "Denoising of ultrasound images affected by combined speckle and Gaussian noise", IET Image Processing, vol. 12, no. 12, pp. 2346-2351, 2018.

[89] L. Ghorbanzadeh, A. Torshabi, J. Nabipour and M. Arbatan, "Development of a Synthetic Adaptive Neuro-Fuzzy Prediction Model for Tumor Motion Tracking in External Radiotherapy by Evaluating Various Data Clustering Algorithms", Technology in Cancer Research & Treatment, vol. 15, no. 2, pp. 334-347, 2015.

[90] D. Ferreira, A. Nordberg and E. Westman, "Biological subtypes of Alzheimer disease", Neurology, vol. 94, no. 10, pp. 436-448, 2020.

VITA

MEHDI SHOJAIE

EDUCATION

2017-2022	Ph.D., Electrical & Computer Engineering Florida International University Miami, Florida
2010-2012	M.S., Electrical & Computer Engineering Sharif University of Technology Tehran, Iran
2005-2010	B.S., Electrical & Computer Engineering Isfahan University of Technology Isfahan, Iran

PUBLICATIONS, PRESENTATIONS, AND PATENTS

M. Shojaie, M. Cabrerizo, S. T. DeKosky, D. E. Vaillancourt, D. Loewenstein, R. Duara, and M. Adjouadi, "A transfer learning approach based on gradient boosting machine for diagnosis of Alzheimer's disease," Frontiers in Aging Neuroscience, vol. 14, 2022.

M. Shojaie, S. Tabarestani, M. Cabrerizo, ST. DeKosky, DE. Vaillancourt, D. Loewenstein, R. Duara and M. Adjouadi, "PET Imaging of Tau Pathology and Amyloid-β, and MRI for Alzheimer's Disease Feature Fusion and Multimodal Classification", Journal of Alzheimer's Disease, vol. 84, no. 4, pp. 1497-1514, 2021.

M. Aghili, S. Tabarestani, C. Freytes, M. Shojaie, M. Cabrerizo, A. Barreto, N. Rishe et al. "Prediction modeling of Alzheimer's disease and its prodromal stages from multimodal data with missing values," International Journal of Medical and Health Sciences 13, no. 2, 36-40, 2019.

S. Tabarestani, M. Aghili, M. Shojaie, C. Freytes, M. Cabrerizo, A. Barreto, N. Rishe, R. E. Curiel,

D. Loewenstein, R. Duara, and M. Adjouadi, "Longitudinal prediction modeling of alzheimer disease using recurrent neural networks," 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2019.

M. Shojaie, N. Elsayad, H. Moradisizkoohi and O. A. Mohammed, "Design and Experimental Verification of a High-Voltage Series-Stacked GaN eHEMT Module for Electric Vehicle Applications," in IEEE Transactions on Transportation Electrification, vol. 5, no. 1, pp. 31-47, March 2019.

H. Moradisizkoohi, N. Elsayad, M. Shojaie, and O. A. Mohammed, “PWM Plus phase-shift- modulated three-port three-level soft-switching converter using GAN switches for photovoltaic applications,” IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 7, no. 2, pp. 636–652, 2019.

S. Tabarestani, M. Aghili, M. Shojaie, C. Freytes and M. Adjouadi, "Profile-Specific Regression Model for Progression Prediction of Alzheimer's Disease Using Longitudinal Data," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1353- 1357.

M. Shojaie, N. Elsayad and O. A. Mohammed, "Design of an all-GaN bidirectional DC-DC converter for medium voltage DC ship power systems using series-stacked GaN modules," 2018 IEEE Applied Power Electronics Conference and Exposition (APEC), San Antonio, TX, USA, pp.2155-2161, 2018.

M. Shojaie and O. A. Mohammed, “A multi-input DC-DC converter with AC-DC PFC Buck-boost stage for Hybrid Energy Storage Systems,” SoutheastCon 2018, 2018.

M. Shojaie, N. Elsayad, S. Tabarestani, and O. A. Mohammed, "A Bidirectional Buck-boost Converter Using 1.3kV Series-Stacked GaN E-HEMT Modules for Electric Vehicle Charging Application," 2018 IEEE 6th workshop on Wide Bandgap Power Devices and Applications (WiPDA), Atlanta, GA, USA, pp. 82-87, 2018

Nour Elsayad, Mehdi Shojaie, O.A. Mohammed, “Multi-Drain Gallium-Nitride Module with Multiple Voltage Ratings”, US Patent App. 15/979,807, 2019.

Nour Elsayad, Mehdi Shojaie, O.A. Mohammed, “Power Electronic Building Block Using Series- Stacked Gallium-Nitride HEMTs”, US Patent App, 15/979,892, 2019.