Florida International University

FIU Digital Commons

FIU Electronic Theses and Dissertations | University Graduate School

11-12-2021

Deep Learning for Multiclass Classification, Predictive Modeling and Segmentation of Disease Prone Regions in Alzheimer's Disease

Maryamossadat Aghili
Florida International University, maghi001@fiu.edu

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

 Part of the Computer Sciences Commons

Recommended Citation

Aghili, Maryamossadat, "Deep Learning for Multiclass Classification, Predictive Modeling and Segmentation of Disease Prone Regions in Alzheimer's Disease" (2021). *FIU Electronic Theses and Dissertations*. 4843.
https://digitalcommons.fiu.edu/etd/4843

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

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

DEEP LEARNING FOR MULTICLASS CLASSIFICATION, PREDICTIVE MODELING AND SEGMENTATION OF DISEASE PRONE REGIONS IN ALZHEIMER'S DISEASE

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

by

Maryamossadat Aghili

2021

To: Dean John Volakis
College of Engineering and Computing

This dissertation, written by Maryamossadat Aghili, and entitled Deep Learning for Multiclass Classification, Predictive Modeling and Segmentation of Disease Prone Regions in 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.

Naphtali Rishe

Mercedes Cabrerizo

Leonardo Bobadilla

Monique s. Ross

David Loewenstein

Malek Adjouadi, Major Professor

Date of Defense: November 12, 2021

The dissertation of Maryamossadat Aghili is approved.

Dean John Volakis
College of Engineering and Computing

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

Florida International University, 2021

© Copyright 2021 by Maryamossadat Aghili
All rights reserved.

DEDICATION

To my mother, a strong soul that taught me how to believe in myself, and to my supportive father who showed me how to be passionate about life.

Without their endless love, support and encouragement, the completion of this work would not have been possible.

ACKNOWLEDGMENTS

First and foremost, I sincerely want to express my gratitude and appreciation to my advisor, Dr. Malek Adjouadi, who has not only been a valuable guide and mentor in my academic life but also taught me great lessons on life for me to contemplate and emulate. I am deeply thankful to all the people who have provided intellectual contributions and motivational support to this dissertation. I have no words to describe all their help and support. Hence, a simple acknowledgment - but a sincere thanks from the bottom of my heart. Each and everyone acknowledged below has taught me something meaningful. My joint advisor and committee members: Dr. Naphtali Rishe, Dr. Leonardo Bobadilla, Dr. Monique S. Ross, Dr. Mercedes Cabrerizo, Dr. David Loewenstein

My family and friends, especially my parents and my husband Dr. Amin Jadidi.

ABSTRACT OF THE DISSERTATION

DEEP LEARNING FOR MULTICLASS CLASSIFICATION, PREDICTIVE MODELING AND SEGMENTATION OF DISEASE PRONE REGIONS IN ALZHEIMER'S DISEASE

by

Maryamossadat Aghili

Florida International University, 2021

Miami, Florida

Professor Malek Adjouadi, Major Professor

One of the challenges facing accurate diagnosis and prognosis of Alzheimer's Disease (AD) is identifying the subtle changes that define the early onset of the disease. This dissertation investigates three of the main challenges confronted when such subtle changes are to be identified in the most meaningful way. These are (1) the missing data challenge, (2) longitudinal modeling of disease progression, and (3) the segmentation and volumetric calculation of disease-prone brain areas in medical images. The scarcity of sufficient data compounded by the missing data challenge in many longitudinal samples exacerbates the problem as we seek statistical meaningfulness in multiclass classification and regression analysis. Although there are many participants in the AD Neuroimaging Initiative (ADNI) study, many of the observations have a lot of missing features which often lead to the exclusion of potentially valuable data points that could add significant meaning in many ongoing experiments. Motivated by the necessity of examining all participants, even those with missing tests or imaging modalities, multiple techniques of handling missing data in this domain have been explored. Specific attention was drawn to the Gradient Boosting (GB) algorithm which has an inherent capability of addressing missing values. Prior to applying state-of-the-art classifiers such as Support Vector Ma-

chine (SVM) and Random Forest (RF), the impact of imputing data in common datasets with numerical techniques has been also investigated and compared with the GB algorithm. Furthermore, to discriminate AD subjects from healthy control individuals, and Mild Cognitive Impairment (MCI), longitudinal multimodal heterogeneous data was modeled using recurring neural networks (RNNs). In the segmentation and volumetric calculation challenge, this dissertation places its focus on one of the most relevant disease-prone areas in many neurological and neurodegenerative diseases, the hippocampus region. Changes in hippocampus shape and volume are considered significant biomarkers for AD diagnosis and prognosis. Thus, a two-stage model based on integrating the Vision Transformer and Convolutional Neural Network (CNN) is developed to automatically locate, segment, and estimate the hippocampus volume from the brain 3D MRI. The proposed architecture was trained and tested on a dataset containing 195 brain MRIs from the 2019 Medical Segmentation Decathlon Challenge against the manually segmented regions provided therein and was deployed on 326 MRI from our own data collected through Mount Sinai Medical Center as part of the 1Florida Alzheimer Disease Research Center (ADRC).

TABLE OF CONTENTS

CHAPTER PAGE

1. INTRODUCTION 1
1.1 Background 1
1.2 Motivation and Problem Statement 2
1.3 Research Objectives and Contributions 6
1.4 Summary and Roadmap 8

2. LITERATURE REVIEW 10
2.1 Background 10
2.2 Related Work on AD Classification 11
2.2.1 Alzheimer Disease vs Cognitively Normal 11
2.2.2 Alzheimer Disease, Mild Cognitive Impairment and Cognitively Normal 13
2.3 Related Work on Patient Sequential Data Modeling 14
2.4 Related Work on Application of Deep Learning for AD Diagnosis 17
2.5 Brain Segmentation 19

3. ADDRESSING THE CHALLENGES OF INCOMPLETE MULTIMODAL DATASETS FOR ALZHEIMER'S DIAGNOSIS 20
3.1 Introduction 20
3.2 Dataset and Preprocessing 24
3.3 Experimental Methods 26
3.3.1 Random Missing Data Handling 32
3.3.2 Original Missing Data Handling 34
3.4 Retrospective 39

4. PREDICTIVE MODELING OF LONGITUDINAL DATA FOR ALZHEIMER'S DISEASE DIAGNOSIS USING RNNS 46
4.1 Introduction 46
4.2 Dataset 47
4.3 Models 49
4.3.1 Long Short Term Memory Unit (LSTM) 50
4.3.2 Gated Recurrent Unit (GRU) 51
4.3.3 Proposed Model 51
4.4 Experiments 54

5. DEEP LEARNING BASED SEGMENTATION MODELS FOR ALZHEIMER DISEASE DIAGNOSIS 58
5.1 Introduction 58
5.2 Related work on Hippocampus Segmentation 61
5.2.1 Locating the Hippocampus in Brain 3D MRI 62

5.2.2 Hippocampus Segmentation 63
5.2.3 CNN Models for AD Diagnosis 65
5.3 Methods and Network Architecture Design 67
5.3.1 Loss Function 74
5.4 Evaluation Metrics 76
5.4.1 Optimizer Algorithm 77
5.4.2 Activation Function 78
5.5 Dataset 79
5.6 Experiments and Results 80
5.7 Discussion 84

6. CONCLUDING REMARKS AND FUTURE WORK 89
6.1 Concluding Remarks 89
6.2 Future Research 91

BIBLIOGRAPHY 95

VITA 128

LIST OF TABLES

TABLE PAGE

3.1 Biomarkers used in the study [RM14, JVW+11] 23

3.2 Comparison Of (A) Random Forest (B) Support Vector Machine (C) Gradient Boosting coupled with five imputation techniques in 4 way classification task 41

3.3 Binary classification of the Control Normal vs Early Mild Cognitive Impairment (CN vs EMCI)* 42

3.4 Binary classification the Early vs Late Mild Cognitive Impairment (EMCI vs LMCI)* 43

3.5 Binary classification the Late Mild Cognitive Impairment vs Alzheimer Disease (LMCI vs AD)* 44

3.6 Binary classification the Control Normal vs Alzheimer Disease (CN vs AD) 45

4.1 Dataset statistics 47

4.2 Model hyperparameters 55

4.3 Performance of the proposed models with three different data arrangements in classification of ADNI subjects between two classes of AD vs NC. Best results for each data arrangement are underlined, and the best overall results of each column are in bold. 56

4.4 Performance of the proposed models with three different data arrangements in classification of ADNI subjects between two classes of AD vs MCI. Best results for each data arrangement are underlined, and the best overall results of each column are in bold. 56

4.5 Performance of the proposed models with three different data arrangements in classification of ADNI subjects between two classes of NC vs MCI. Best results for each data arrangement are underlined, and the best overall results of each column are in bold. 57

5.1 Performance of UNet model paired with different feature extraction backbones. When a reference is not available, * means implemented by our group 83

5.2 Performance of UNet++ model paired with different feature extraction backbones. When a reference is not available, * means implemented by our group 83

5.3 Performance of MANet model paired with different feature extraction backbones. All combination have been proposed by our group 84

5.4 The Performance of different segmentation methods over Dechathlon dataset. *Results of best performing combination of backbone for UNet,UNet++ and MANet models are illustrated. 84

LIST OF FIGURES

FIGURE PAGE

3.1 Effects of imputation methods coupled with Random Forest classifiers at different degrees of random missing values on 4-way (i.e. multiclass) classification (CN, EMCI, LMCI, AD) of the subjects 32

3.2 Effects of imputation methods coupled with Support Vector Machine classifiers at different degrees of random missing values on 4-way (i.e. multiclass) classification (CN, EMCI, LMCI, AD) of the subjects 33

3.3 Effects of imputation methods coupled with Gradient Boosting classifiers at different degrees of random missing values on 4-way (i.e. multiclass) classification (CN, EMCI, LMCI, AD) of the subjects . . 33

3.4 Performance of Random Forest on real missing data (blockwise missing) coupled with different imputation techniques 35

3.5 Performance of Support Vector Machine on real missing data (blockwise missing) coupled with different imputation techniques 35

3.6 Performance of Gradient Boosting on real missing data (blockwise missing) coupled with different imputation techniques 36

3.7 Receiver Operating Characteristic of Random Forest for ADNI dataset 37

3.8 Receiver Operating Characteristic of Support Vector Machine for ADNI dataset 37

3.9 Receiver Operating Characteristic of Gradient Boosting for ADNI dataset 38

4.1 Sample datapoint curation 49

4.2 An RNN network with LSTM cell with (k+1) time points. 50

4.3 The structure of LSTM and GRU cells 51

4.4 Data arrangement for RNN model 53

4.5 Schematic of RNN model and how input data feed to the model 54

5.1 The two-stage proposed pipeline for hippocampus segmentation in MR images 71

5.2 Detailed illustration of the proposed architecture 75

5.3 Illustration of model segmentation against FreeSurfer and manual segmentation masks in different subjects has shown in left and right columns respectively. Each subject hippocampus has been visualized in axial, sagittal and coronal view 85

5.4 Performance comparison of top performing models integrated in proposed two-stage pipeline on hippocampus segmentation task. 86

CHAPTER 1

INTRODUCTION

1.1 Background

Alzheimer's Disease (AD) is one of the most prevailing cause of dementia. AD is a progressive and irreversible neurodegenerative disorder that leads to loss of neural cells, loss of memory, language impairment, and other cognitive and functional disabilities.

The National Institute on Aging (NIA) and the Alzheimer's Association provided guidelines for AD diagnosis. They have defined three main stages for dementia subjects including pre-clinical AD, mild cognitive impairment (MCI), and Alzheimer's disease (AD). While subjects in the pre-clinical AD stage do not show significant symptoms of cognitive impairment, they have noticeable changes in their brain structure. Subjects in the MCI stage are divided into early MCI (EMCI) and late MCI (LMCI). Subjects in this group still can perform their daily tasks, but they exhibit some degree of cognitive impairment. Even though there is a high convergence rate from MCI to AD stage, this progress can be halted or even reverted for some subjects [SAB+11, ADD+11, JJAK+11].

AD accounts for 60 to 80 percent of dementia cases [BY11] and it is one of the leading causes of death in the United States. It is a very common disorder that afflicts more than 3 million cases per year in the United States alone. As people are living longer and aging being the main risk factor, the rate of patients with AD is growing. In 2020 the Alzheimer's Association announced that without proper prevention techniques, the number of AD subjects will rise to more than 100 million worldwide by 2050. Therefore AD research has gained lots of attention in recent decades to develop effective ways to diagnose this disease in its earliest stage and

to propose or plan for early treatment and intervention protocols. Accurate AD diagnosis and prognosis is of critical importance, especially for (1) early disease detection through more precise delineation of the Early Mild Cognitive Impairment (EMCI) group from the Cognitively Normal (CN) control group, and (2) for discriminating possible converter mild cognitive impaired patients from non-converter subjects [CGT$^+$11a, PM05a].

1.2 Motivation and Problem Statement

Early diagnosis of AD, before reaching an irreversible late stage of the disease, is highly important. New algorithms should be designed to ensure more accurate classification and prediction of measures that contribute to the transition phases of the disease, all in order to plan early for treatment and therapeutic interventions [PM05a, MPG$^+$15a, IMC$^+$17]. However, regardless of the enormous efforts, delineating the early stage of mild cognitive impairment (EMCI) from cognitively normal (CN) controls remains an open research endeavor. Furthermore, available data for AD studies usually experience a high number of missing values due to issues related to subjects missing out on follow-up visits for several personal and health-care related reasons. Having incomplete samples in longitudinal medical studies is a common phenomenon, as many patients may miss some of the tests and imaging modalities at a given time step of the study, miss complete visit(s) within the study's lifespan or simply withdraw because of unforeseen reasons [IMC$^+$17, DFA$^+$09, SLS$^+$14a].

Many studies reported in the literature simply discard subjects with missing modalities and related measures, which results in considerable loss of valuable information, reducing as a consequence the statistical significance of the study with a lesser number of subjects considered. There is great potential that disease diag-

nosis may be improved if the missing parameters could be estimated correctly from the rest of the available data or modalities. In addition, resolving the missing data challenge may lead to a better understanding of the disease and to a more reliable labeling of subjects as Cognitively Normal (CN), Mild Cognitive Impairment (early and late MCI), or dementia (i.e., AD) from baseline to every visit thereafter when new data is collected at each time step. Currently, the majority of the classification algorithms make use of cross-sectional data involving baseline measurements for the diagnosis, without regard to any other time point as a reference to disease progression evaluation and for a more informed decision-making process. To address this shortcoming, recent studies moved toward longitudinal data analysis and proposed new methods to leverage valuable temporal data by considering the inherent correlations of such data [DFA+09, SLS+14a, LHM+10]. Effective mining of AD in longitudinal studies is also a challenging task, owing to its heterogeneous measurements coming from various sources and modalities, varying length of samples, missing modalities and tests, small and imbalanced sample data, to name a few.

Projecting a trajectory to gauge in a meaningful way disease progression has been constrained for a long time due to a lack of sufficient longitudinal data. In recent decades, ADNI has released a relatively large dataset of AD subjects in longitudinal studies that range from 5 to 10 years, an incredible feat which is enabling researchers to focus on progression modeling of the disease.

Alzheimer's Disease progression is commonly evaluated using biomarkers including structural Magnetic Resonance Imaging (MRI), functional Magnetic Resonance Imaging (fMRI), 18-Fluoro-DeoxyGlucose (FDG) Positron Emission Tomography (FDG-PET) imaging, Cerebro-Spinal Fluid (CSF), cognitive examination, and to a lesser extent electroencephalography (EEG) [JPM18, PDHvdF+13, LCD+18]. There are a group of studies that focus on a single biomarker change in the course of a

time while others consider multiple modalities and combine the different biomarkers for diagnosis and prognosis.

With the advent of neuroimaging technologies in the past few decades, several important structural and functional changes in the brain can now be captured leading to enhanced diagnosis of the different neurological and neurodegenerative disorders. The most informative ones for AD diagnosis are brain anatomical structure deformation, deposition of protein outside neurons, and deposition of tau protein inside nerve cells. Hence structural Magnetic Resonance Imaging (sMRI)[1], functional MRI (fMRI), and Positron Emission Tomography (PET) became the three main neuroimaging modalities used for AD diagnosis.

In this dissertation, among the few disease prone regions that are commonly studied, of interest is the Hippocampus region viewed here as one of the most informative early diagnostic biomarkers of AD and many other neurological and neurodegenerative disorders. Early detection of hippocampus shape changes and accurate measurement of hippocampus atrophy through volume measurements greatly support the early detection of MCI subjects. MRI has shown to be a promising non-invasive modality for AD diagnosis via imaging of brain internal structures including the hippocampus. This neuroimaging modality assists experts to understand brain anatomical changes including cortical thinning and brain atrophy related to AD. However experts are required to delineate the region of interest or extract brain features manually for a variety of clinical applications that rely on hippocampus volumetric measurements as means to improve disease diagnosis, estimating disease progression, and assessing subject's response to treatment and therapy, etc. [SAB+19].

[1]sMRI and MRI are used interchangeably in this dissertation

Considering the importance and effectiveness of the MRI modality in the medical field plus the difficulty and tediousness of manual medical image segmentation, automatic segmentation, classification, feature extraction, pattern recognition, and image processing methods have all experienced incredible theoretical and practical advances in the realm of medicine. This line of research aims to create a fully automated pipeline that captures the image, detects and segments the organ of interest, extracts features, classifies the subject and diagnose the disease.

Without any doubt, deep learning has been the most significant breakthrough of the last decade in several research areas including artificial intelligence, computer vision, and natural language processing [CVMG$^+$14a, KSH17, HZRS16, RDS$^+$15]. The fundamental concepts of deep learning such as fully connected layer and back-propagation which date back to the 1980s have gained popularity lately with the advent of Graphical Processing Units (GPU) and increased processing power of computers. The release of several large-scale annotated datasets such as ImageNet (with over 14 million images designed to advance computer vision research)and COCO dataset (large-scale object detection, segmentation, and captioning dataset) was another significant contributor to the development of advanced machine and deep learning methods.

To address the aforementioned challenges and take advantage of deep learning models as they apply to neuroimaging, especially for AD biomarkers detection, diagnosis and prognosis modeling, new techniques for multiclass classification and regression analysis are developed while contending with the missing data challenge and the segmentation challenge of disease prone areas like the hippocampus.

1.3 Research Objectives and Contributions

The main objectives when studying a given disease have always been early diagnosis, regression analysis and disease progression, and the planning of effective treatment plans with the means to gauge pre and post treatment results. A deep understanding of Alzheimer's disease calls for a computing environment that integrates multimodal imaging for structural and functional brain data, neuropsychological testing, genetics and demographic factors, all linked to a database with common evaluations and standardized measures that are amenable for multimodal and multi-site studies and the merging of data across sites. A first part for this kind of research endeavor relates to the analysis of neuroimaging data to be consolidated with cognitive scores and other genetic and demographic variables as we seek optimal classification and regression analysis. The next step is to design algorithms that are amenable to both cross-sectional and longitudinal studies that can overcome the missing data challenge.

The interplay of these diverse measures from existing biomarkers is so complex and in many cases so subtle that, without the advent of new biomarkers, it is quite complicated to correctly model Alzheimer's Disease in its different stages. Complex and diverse forms of disease progression in different patients with different backgrounds, cultures, and medical history, make this problem even more complex. Therefore biomarker discovery and relevant feature extraction, accurately determining the decision boundaries of Cognitively Normal (CN) vs. Early Mild Cognitive Impairment patients (EMCI) , or EMCI vs. Late Mild Cognitive Impairment patients (LMCI) in a multiclass classification scenario, and precisely modeling the progression of the disease are three prospective research areas this dissertation aims to explore. Large neural networks contain a sizeable number of trainable parameters

that in theory can capture the relationships between input-output pairs. However, it is worth mentioning that deep learning based-approaches although proven to be very powerful in several fields, their power is drawn from the availability of a huge amount of data (in the millions and up), which is not particularly true for Alzheimer's studies (currently involving fewer than two thousand individuals). However, when the most relevant features are extracted, high accuracy can be attained even for smaller samples.

Along this line of thought, the missing data challenge is thus an important problem that first part of this dissertation addresses for two very important reasons: (1) to provide added statistical significance with more samples kept in the analysis, and (2) gauge the effect and merit in their inclusion through the accuracy of the results obtained.

On the other research front, while one of the most significant biomarkers for early detection of this disease is the hippocampus region; correct segmentation and volume estimation of this region of the brain can be of remarkable importance in disease diagnosis. Therefore, the second part of this research focuses on hippocampus volume estimation and segmentation with the application of deep learning techniques. In this research endeavor, a two-stage model is proposed for localization and segmentation of the hippocampus in 3D MRI. In the localization module, a heuristic model estimates the location of the hippocampus and crops a cube-like structure surrounding that region. The second module which is a novel segmentation model is composed of Vision Transformers and UNET architecture with a tailored loss function for imbalanced segmentation. By integrating these two stages, a fully automated diagnostic pipeline is designed with minimal need for expert intervention and for prior preprocessing. More importantly, as another contribution of this dissertation the proposed algorithm is tested on two different datasets that prove its

easy deployment and the validity of the results obtained which are compared to manual segmentation and to a popular software suite called FreeSurfer commonly used for brain segmentation.

1.4 Summary and Roadmap

The rest of this dissertation is structured in the following manner: Chapter 2 is a literature review that pertains to AD classification and the difficult tasks that relate to early detection and the potential delineation of early mild cognitive impairment (EMCI) to cognitively normal (CN) subjects in a multiclass classification scenario. This chapter also delves in to related work on sequential data modeling, deep learning and brain segmentation.

In chapter 3, the work on the AD classification problem with missing data is studied, where the problem of drawing boundaries is investigated between four classes of subjects (AD, LMCI, EMCI, NC). In the same study, multiple imputation techniques are explored to define those algorithms that can best assist the machine learning model in the presence of missing values. This Chapter outlines the superiority of the Gradient Boosting algorithm in terms of accuracy, precision, and recall in both cases of missing at random or block-wise missing data.

In chapter 4, the focus is placed on sequential modeling of longitudinal data with deep learning methods, where different variations of the recurrent neural network are employed to assist trajectory prediction of the subjects. Different alternatives at overcoming the missing data challenge are deployed and the results for each are provided.

Chapter 5 presents an advanced two-stage deep learning-based framework, which is a combination of Vision Transformer (ViT) and Convolutional Neural Network

(CNN) for hippocampus segmentation. In this chapter, the effectiveness of CNN models is demonstrated for object localisation and image segmentation. Two types of experiments are conducted including (1) the Medical Segmentation Decathlon Challenge of 2019 involving 194 (3D) volumes for training and validation through cross validation technique and the rest, the 65 (3D) volumes for testing; and (2) our own Mount Sinai data with the 1Florida Alzheimer's Disease Research Center (ADRC) data, consisting of 326 MRI volumes with a comparison made with the FreeSurfer version 6.0 results

Finally, in chapter 6, we conclude the work of this dissertation and discuss the future work along this line of research in terms of both imputation techniques and brain segmentation methods.

CHAPTER 2

LITERATURE REVIEW

2.1 Background

Early diagnosis of AD, given the knowledge that the disease may have started a decade or so earlier before its first symptoms appear, is critically important in planning for patient's treatment and slowing progression of the disease. Many researchers have conducted experiments to explore and understand the underlying factors of this disease. Great attention has been devoted to biomarker discovery, feature extraction, classification and regression analysis using various sources including neuroimaging modalities such as Magnetic Resonance Imaging (MRI)[DBS+11], Positron Emission Tomography (PET)[NRKL10], and functional MRI (fMRI)[GSRM04]. Other relevant biomarker sources include cerebrospinal fluid (CSF)[DBS+11, NRKL10, GSRM04, ZWZ+11], electroencephalographic (EEG) rhythms [MCL+15a], genetic tests, demographic information and neuropsychological testing [TYKM05] including Montreal Cognitive Assessment (MoCA), Mini-Mental State Exam (MMSE), Clinical Dementia Rating (CDR) Scale, and the Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS-Cog), among the most relevant.

With the advent of machine learning and deep learning in the medical field and the biosciences in the recent decade, various algorithms and techniques have been proposed for analyzing different biomarkers and classification of subjects into the three common groups of Cognitively Normal (CN) controls, Mild Cognitive Impairment (MCI) and Alzheimer Disease (AD). However, the number of studies is limited when it comes to the subdivision of the MCI subjects into the two subgroups of Early MCI (EMCI) and Late MCI (LMCI) due to reasons such as limited data from the targeted groups, complex form of progression in the different subject groups, miss-

ing biomarkers, non-unified availability of features and modalities from each subject, challenges of working with neuroimaging data, difficulties of brain anatomical region segmentation and feature extraction, and most of all lack of extensive longitudinal studies, which are extremely difficult to collect.

To better understand current state-of-the-art methods and algorithms proposed in this domain, this chapter provides a brief review and summary of some relevant research and related articles that address different challenges of AD diagnosis and prognosis with the help of machine learning and deep learning techniques.

2.2 Related Work on AD Classification

Automated classification of AD subgroups and Cognitively Normal (CN) subjects with the focus on early detection is one of the first and most known challenges in the development of classification models. Many algorithms including Support Vector Machine (SVM), Artificial Neural Network (ANN), Multi-Layer Perceptron, Decision Tree, Random Forest, Bagging and Boosting, etc. have been explored for classification modeling of AD and its prodromal stages . These algorithms were incorporated with a variety of feature selections and dimensionality reduction methods to accurately draw a boundary between the different groups of subjects. While some studies focus on a single modality of biomarkers, there have been others that combine several recording modalities and model the high dimensional feature space. Some of the classification algorithms for the AD diagnosis are discussed below.

2.2.1 Alzheimer Disease vs Cognitively Normal

SVM is one of the techniques that have been widely studied in this domain by multiple researchers due to its power on the correct modeling of multi-dimensional

spaces. Vemuri et al. have combined the sMRI features with demographic and genetic scans of each subject and applied SVM model to draw a boundary between AD and CN subjects [VGS+08]. Magnin et al. applied a feature selection method followed by an SVM model trained on the features of whole brain anatomical MRI to classify subjects into two main groups [MMK+09]. Gaussian Mixture Models (GMM) have been used to reduce the number of features extracted from functional MRI, after that the SVM model has been trained on the selected feature set to model the subjects' space [SGR+10]. Sparse Inverse Covariance Estimation (SICE) method combined with SVM have been utilized to discriminate subjects by using their PET and structural MRI features in [OMÁI+15]. Bi et al, have trained multiple SVM models on random sub-samples of features, captured from resting-state functional MRI modalities. They have ensembled the SVM predictions to generate the label for every sample. While they have improved the classification accuracy with this model, they have not exploited the other modalities' contribution in the classification of the subjects [BXL+18].

One of the other widely employed techniques for AD subject classification is Artificial Neural Network (ANN). Several researchers have combined different feature selection techniques from various modalities and trained different ANN models using different classification tasks. Principle component analysis has been applied to voxel-based morphometry (VBM) biomarkers to select the most informative features. Afterward, an ANN model is trained to learn the feature space to automatically detect gray matter loss in AD subjects, thus classifying the subjects accordingly [HYJW08].

Several other studies tackled different types of binary classification tasks using ANN [SA11], Logistic Regression [GKH+90, JVW+14], and Random Forest[LXC+17] to name a few.

2.2.2 Alzheimer Disease, Mild Cognitive Impairment and Cognitively Normal

Earlier attempts in automatic classification of AD subjects focus more on the classification of subjects into two groups of AD and non-AD in a binary type of classification. However, the more recent studies have devoted great efforts to the early detection of AD, with few extending such efforts into the more challenging multiclass classification. Studies that try to develop interesting algorithms for delineating the prodromal stage of mild cognitive impairment (MCI) show varied and competing classification results [SLS+14a, LHM+10, Jag18].

In [GCC+09], hippocampal shape features have been estimated by spherical harmonics (SPHARM) coefficients algorithm to model the shape of the hippocampal region. These shape features have been fed to an SVM algorithm to model three classes of subjects (AD, MCI, NC), with NC defining the Non-converter group. SVM with kernel function has been utilized in several research papers for multiclass classification of the subjects in a similar manner [LB14].

Gorji et al. have presented a Pseudo Zernike moments to extract discriminate information from subjects' MRI. They have used ANN to delineate a boundary between three groups of AD, MCI, and NC [GH15].

Random Forest (RF) algorithm is another widely studied algorithm for AD subject detection. In [GAH+13a] several features from four modalities of CSF biomarker measures, regional MRI volumes, categorical genetic information, and voxel-based FDG-PET signal intensities have been combined to train a random forest algorithm for AD classification. The study of Wang et al. applied SVM, partial least square, and Random forest to classify the subjects based on 3 sets of biomarkers from 1.5 T MRI, FDG-PET and florbetapir-PET modalities [WCY+16].

This line of research has evolved during the last few decades and various biomarkers from different modalities including CSF, Genetic, MRI, PET, fMRI, etc. have been deployed to train a wide range of machine learning algorithms for better automatic classification of the subjects. Comprehensive reviews on this topic have been provided by [SCQ17, TRK+20].

Most researchers agree that the current accuracy that is achieved, especially for classifying EMCI vs. CN groups remains below an acceptable standard for the medical field in light of the irreversible nature of the disease. This classification error is of course further influenced in the negative when performing multiclass classification (assuming all groups as it should be) rather than binary classification (when only two classes are assumed at a time) and when contending with the missing data challenge. Lack of sufficient data with complete samples for all the subjects considered in a study whether cross-sectional or more significantly longitudinal is an inherent problem of any clinical trial.

2.3 Related Work on Patient Sequential Data Modeling

Until recently, most of the research on AD topics were mainly focused on classifying the subjects between groups (AD vs CN, AD vs MCI vs CN, NC vs EMCI vs LMCI vs CN, etc.) based on cross-sectional data points. However, these approaches lack enough information to predict a future state of the subject to take proactive steps in treatment planning. While early diagnosis is of critical importance for slowing down the progression of the disease through early intervention, researchers have started to track and monitor subjects in time and create a longitudinal dataset to better analyze the biomarkers and project the progression of the disease in different groups of subjects. The release of the new longitudinal AD dataset by ADNI opens up many

opportunities for researchers to leverage temporal information from longitudinal data and model AD progression. Several classification and regression models have been proposed to model this longitudinal data.

Duchsesne et al. have applied principal component analysis on MRI-based features (local volume changes, intensities). They combined MRI features with baseline MMSE, gender, education level, and age of the subjects and trained a linear regression model to predict the MMSE score of the subjects in the following year [DCG+09]. Longitudinal changes of the AD subjects' cortical thickness and its effect on the development of AD have been studied in [LWW+12]. Where they exploited the longitudinal data to classify the subjects between 4 groups of Stable-MCI and Progressive-MCI, AD, and CN. Cortical thickness, cortex thinning speed, and thickness changes of different regions of interest are some of the important features they have utilized for training an SVM classifier. Jack et al, studied 3 consecutive tau PET, MRI, and clinical assessments of 126 individuals spread in 3 categories of 1) cognitively unimpaired with normal amyloid, 2) cognitively impaired 3) cognitively unimpaired with abnormal amyloid. They have used Ridge and Lasso regression models to depict the trajectory of the disease in all 3 subject groups [JJWS+18].

Unlike many models which investigate the trajectory of continuous brain clinical variables in the future separately, [ZSI+12a] focused on modeling multiple variables together. In this way, the authors exploited the intrinsic correlation information among different variables. They also captured features from multiple modalities instead of focusing on a single modality. They proposed a multitask model to project the Mini-Mental State Examination (MMSE), Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog), and categorical variable (with the value of 'AD', 'MCI' or 'CN'), by using CSF data, baseline MRI, and FDG-PET.
newpage

Zhu et al. proposed a multitask convex and non-convex fused group lasso regression for modeling the temporal relationship between multiple future time points to accurately predict the cognitive scores of the patient in the future [ZYLY11a]. Even though the temporal dependency assumption cannot be guaranteed in reality [HJG+16a]. In 2016, Moradi et al. studied the relationships between AD-related structural atrophy within the brain MRI with Rey Auditory Verbal Learning Test (RAVLT) cognitive measures over 3 years. They utilized an elastic net algorithm for modeling the atrophy in MRI [MHH+17]. In a similar attempt, Wang et al. presented a multi-layer multi-target regression model for clinical multivariate prediction in AD [WZL+18]. This model can simultaneously handle the nonlinear relationship between MRI neuroimaging biomarkers and the cognitive assessment scores. They employed matrix elastic nets to investigate the inter-correlations between multiple test scores. Non-smooth L2, L1-norm loss function is shown to add robustness to their proposed multi-target prediction model.

To address the sparsity in the data and model the cognitive scores in five future time points in longitudinal data, Huang et al. in [HJG+16b] proposed the soft-split sparse regression-based random forest (RF) model. Focusing on the volumetric features of the MRI regions of interest (ROI), they have defined the most discriminate regions along with the future score of the patients only based on the baseline data. Although they provided predictions for multiple future time points, they have relied on the features of the prior time points for every prediction rather than only based on the baseline data. This means that the model cannot predict the trend of the patient's progression in time. Moreover, they have used a single modality (MRI) to model the relationship between patients and disease progression and discarded many other informative biomarkers from other modalities and demographic data.

Some other studies have exploited regression modeling on the longitudinal data

for diagnosis and prognosis purposes [EBP+18, SLS+16, MCL+15b]. However, none of them have applied deep learning techniques for regression modeling of the subjects.

2.4 Related Work on Application of Deep Learning for AD Diagnosis

Early diagnosis of AD plays an important role in slowing down the disease's progress. This diagnosis is particularly based on features that are extracted from brain neuroimages. The features are mainly captured from variations of anatomical brain structures such as hippocampus shape, hippocampus volume, cortical thickness, ventricles size, and brain volume. Automatic discrimination of brain images that are coming from various modalities can be performed in two major ways. 1) the whole brain gets analyzed and global features extracted from the whole brain [HAKEB16] 2) Some specific anatomical regions of the brain get investigated which are proven to have the most discriminating features. Regions such as hippocampus structure that its atrophy is known to be highly correlated with AD onset [ZLA+17].

Several earlier studies for brain segmentation and feature extraction either require human annotation or depend on third-party software [BRSM10]. In both cases, the process is very time-consuming, inefficient, and error-prone. Brain registration (affine or rigid) is still a resource-intensive and slow process. Therefore algorithms and models that do not need human intervention and can work on neuroimaging directly are in high demand.

Billones et al. have employed a modified VGG network to classify subjects based on sMRI images. The trained 2D convolutional network processed each slice of the 3D MRI data separately. They mainly picked 20 central slices of each image

[BDHN16]. Liu et al. in [LZAS17] have defined 50 discriminative anatomical brain landmarks in a data-driven way. Then they have captured multiple image patches from each of the landmarks. Using those features they trained a multi-task multi-channel convolutional neural network for AD diagnosis task to assign subjects to four groups of NC, stable MCI (sMCI) a progressive MCI (pMCI), and AD. To identify the hippocampus location in the brain MRI, Hajiesmaeili et al have proposed an algorithm that performs 3D skull stripping and extracts the brain volume first followed by a distance estimation from the first slice of the brain to the first slice where the hippocampus appears in all 3 views of coronal, sagittal, and axial [HA17]. This algorithm is particularly beneficial if only patches of the hippocampus should be processed instead of the whole brain image.

While many methods for brain automatic feature extraction and subject classification have been proposed, they are mainly limited by one drawback: independent feature engineering and disease modeling. These approaches usually result in sub-optimal performance. Therefore Cao et al. have addressed this issue by proposing a multi-task deep learning model which jointly segments the hippocampus and also models the clinical regression score. Their model does not need prior time-consuming non-linear registration, though they align the MRI images onto a standard template before inputting to the network. After alignment, the model approximately locates the hippocampus based on the template atlas and extracts image patches from those regions. They used those patches to train a network composed of two sub-modules. The first module is a UNET with Dice-loss for hippocampus segmentation and the second network is a CNN with mean squared error loss function for regression modeling of the clinical score. These two modules share some parameters for better utilizing the inherent association between two tasks of segmentation and regression [CLZ+18]. A similar approach has been adopted in [LLY+20] where they have de-

veloped a multi-task deep learning model to jointly segment the hippocampus and classifies the AD subjects. To extract features from hippocampus patches, they employed 3D DenseNet.

2.5 Brain Segmentation

Segmentation has been a critical task in medical image analysis as it allows for the quantification of key anatomical regions and to assess observed changes in cross-sectional or longitudinal studies [GVOV+16]. Segmentation can also help in comparative studies using multimodal neuroimaging and multiclass classification. In neuroimaging, several methods have been proposed for various tasks including brain extraction, anatomical ROIs segmentation, White Matter Lesion (WML) segmentation, brain tumor segmentation, etc. Each of these tasks has its specific challenges. With the recent advances in image processing and computer vision and with the advent of CNNs, automated segmentation of medical images have also made significant progress. Moreover, these segmentation techniques through other imaging or machine learning approaches could extend to other organs like lungs or liver, as our research group has reported in [ETA+20, GGB+14]

CHAPTER 3

ADDRESSING THE CHALLENGES OF INCOMPLETE MULTIMODAL DATASETS FOR ALZHEIMER'S DIAGNOSIS

3.1 Introduction

Early diagnosis, prior to reaching the stage of irreversible disease changes in the brain, allows for the planning of early treatment and therapeutic interventions, and plays a significant role in providing subject-specific care, predicting disease progression, and gauging the rate of decline and severity of impairment [MPG+15a, IMC+17, DFA+09]. The more recent studies have devoted great efforts for early detection of AD by developing algorithms for delineating the prodromal stage of mild cognitive impairment (MCI) with varied and competing classification results [SLS+14a, LHM+10, Jag18]. Most researchers agree that the current accuracy that is achieved, especially for classifying EMCI vs. CN groups remains below an acceptable standard for the medical field in light of the irreversible nature of the disease. This error in classification is of course further influenced in the negative when performing multiclass classification (assuming all groups as it should be) rather than binary classification (when only two classes are assumed at a time) and when contending with the missing data challenge. Lack of sufficient data with complete samples for all the subjects considered in a study whether cross-sectional or more significantly longitudinal is an inherent problem of any clinical trial. This is largely due to the fact that many patients may miss some of tests at different intervals throughout a study. Generally, missing values occur for a variety of reasons, including subjects that stop attending the study or miss appointments, subjects that completely drop out from the study, data with insufficient or incompatible resolutions, image corruption, budget limitation, etc. [TCS+01][LJ12]. Many algorithms simply discard subjects

with missing modalities from further experiments or, in the simplest case, they just replace them with zero values or with a mean average of the attribute, which still results in a loss of valuable information. This challenge of missing data continues to hinder the needed progress for understanding this complex brain disorder [Jag13]. Accuracy in AD diagnosis and prognosis could be improved if the missing parameters can be more precisely estimated from the rest of the available data or modalities or a more reliable technique rather than mean substitution devised [BHR+16]. However, different data modalities often have nonlinear and complicated correlations, which impedes the prospects for correct estimation.

These difficult issues have led to a new line of research that focuses on developing more realistic and more sophisticated techniques to substitute the incomplete samples. This line of research is generally divided into two main approaches: the first approach attempts to synthesize missing modalities from remaining modalities with the help of various techniques that include maximum mean discrepancy based multiple kernel learning [ZTA+17], missing modalities imputation via cascaded residual autoencoder [TLZJ17], 3D convolutional neural networks[PM15] and generative adversarial networks (GAN) [NTL+17, STC+18], However Cohen and his colleagues have pointed out that synthesized medical images may result in mis-diagnosis due to the distribution matching losses that arise from the process of matching an image in the input domain to an image in the target domain while preserving the source distribution [CLH18]. The second approach attempts to impute missing values by applying various numerical techniques such as simple Mean substitution[1], Mode and K-Nearest Neighbor (KNN) impute [CPV+15, LGH12, HJG+16c] . Authors in [STC+18, RSW+15] have extracted a complete subset of features from the actual

[1]As the data is normalized around the center in this study Mean substitution in this case is the same as zero fill.

dataset and synthesized the missing values randomly to analyze the power of some imputation methods, but they have not determined the performance of the algorithms on different patterns of missing values in real incomplete datasets which may actually have completely different patterns from those that were randomly synthesized. They also overlooked the fact that some of the proposed imputation methods assume that the data have a Gaussian distribution, which may not be the case for every dataset. These approaches do not address the block-wise missing patterns of data in the relatively small size dataset of AD. Therefore, to the best of our knowledge, none of the research studies so far have done a comparative study on effects of existing imputation techniques on a block-wise missing dataset of Alzheimer while incorporating a huge sample size from various modalities (PET, MRI, Cognitive Test, CSF) to check the effects of large size data on imputation task. As an additional endeavor, we have also considered the challenging task of multi class classification of the ADNI dataset with a high number of missing points. Moreover, there are a number of new imputation techniques which have never been deeply studied within this scope of work.

Considering the importance of the early detection of the prodromal stage of AD, the first objective of this chapter is to present the classification power of Gradient Boosting (GB) technique on a four-way classification of Cognitively Normal controls (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (MCI) and Alzheimer Disease (AD) subjects on a large multimodal heterogeneous dataset pulled from various cites with missing data, which include ADNI1, ADNIII and ADNIGO. The recent release of ADNI data which discriminate Early and Late MCI patients motivated us to focus on multi class classification between four groups of subjects rather than binary classification of three groups of subjects which has been studied for several years and lacks the generalization power when it comes to

Table 3.1: Biomarkers used in the study [RM14, JVW+11]

Source	Features
Cognitive Test	EcogPtMem, EcogPtLang, EcogPtVisspat, EcogPtPlan, EcogPtOrgan, EcogPtDivatt, EcogPtTotal, EcogSPMem, EcogSPLang, EcogSPVisspat, EcogSPPlan, EcogSPLang,EcogSPOrgan, EcogSPDivatt, EcogSPTotal, FAQ, MOCA, RAVLTforgetting, RAVLTpercforgetting, RAVLTimmediate, RAVLTlearning
MRI	Cognitive TesVentricles, Hippocampus, WholeBrain, Entorhinal, Fusiform, MidTemp, ICV, FLDSTRENG, FSVERSION
PET	FDG, PIB amyloid, AV45 amyloid, CDRSB
Genetic	APOE4
Demographic	AGE, Gender, Education
CSF	Aβ1, t-tau,p-tau

new sample data with no prior diagnosis label. The challenge of discriminating the EMCI group from LMCI has not yet been well studied due mainly to the absence of certain data in those two classes. The second objective of this chapter is to represent the classification potential of GB when it applies to incomplete data sets. While SVM and RF are unable to feed with incomplete sample data, GB is capable of handling missing values with no need of further preprocessing.

Moreover the dependency of the various state of the art imputation techniques on the patterns of missing data has been described. For this purpose, we investigated the performance of a group of imputation techniques on two separate sets of synthesized incomplete data with random wise missing values and real incomplete data with block-wise missing values. Results reveal the shortcomings of imputation techniques in the real case of block-wise missing data estimation. Despite few papers that attempted to proceed in this direction [CPV+15, JMC+16], to the best of our knowledge, this work is the first one that provides an exhaustive comparative study over real, incomplete heterogeneous multimodal dataset of Alzheimer with four groups of subjects (AD, LMCI, EMCI, CN)

3.2 Dataset and Preprocessing

Data used in the preparation of this article were obtained from the Alzheimer Disease Neuroimaging Initiative (ADNI) database (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 serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment (MCI) and Alzheimer's Disease [LHM$^+$10]. ADNI data is processed with a standard pipeline resulting in a large matrix of patients and their test measurements. Patients are arranged in rows and each test result is ordered as column. In this research endeavor, various groups of biomarkers including CSF, MRI, PET, DTI, Genetics, and neuropsychological tests, which are derived from ADNI database have been used. The detailed list of biomarkers is as provided in the table 3.1. Diagnosis labels are composed of Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI) and Alzheimer's Disease (AD).

The dataset that was considered for this study consists of 1627 subjects (413 CN, 312 EMCI, 565 LMCI and 337 AD) which have been examined during an 11-year period and for every six months. However, many of the subjects have only a couple of visits or time points. There is also diagnosis label for every visit separately, regardless of the earlier diagnosis in the previous visit, but we mainly focus on the baseline data for our experiments. A comprehensive set of 41 biomarkers were selected at the end.

Imputation techniques tend to perform better in huge datasets, with more samples for training the model in search of an optimal classifier. Each patient-visit

was regarded as a separate sample to augment the data size, help the imputation process and augment the prospects in establishing or modeling an optimal classifier that generalizes better. It is worth mentioning that even if the subject status does not change between successive time points, subject-sample vector at various time points tends to be slightly different from each other due to the alteration in imaging and other tests.

In contrast to most studies that handle the problem of missing data, either by excluding the patients with incomplete test results or by restricting the study to a single modality, we tried to solve the problem by introducing an algorithm which is capable of handling the missing values naturally. In a parallel experiment an imputation stage was added to two other classifiers of RF and SVM to check if this incorporation adds any robustness to the classification algorithms. Furthermore, in contrast to the majority of studies which work on a dataset from a specific single source, current ADNIMERGE dataset is pulled from multiple cites which adds heterogeneity to data and makes the classification process more challenging. Pre-analysis of the dataset indicates that on the basis of all the aforementioned biomarkers considered in this study, there was not a single record that did have not have one or more measurements missing. The number of missing values altogether for all patients' samples and throughout all the biomarkers is equivalent to 46% of the entire dataset.

This clearly highlights the challenges researchers face when preforming longitudinal studies while contending with missing data. It also places research groups in the predicament of choosing only those complete datasets albeit ending up with a smaller size dataset and more likely with a lesser number of recording modalities (i.e., measurements), or assume the entire data set but find ways to impute the missing data in the best way possible that would preserve the statistical and clinical meaningfulness of the longitudinal data that took so many efforts and so many years to establish.

Since feature normalization is required for many algorithms, especially in the cases of SVM and K-nearest neighbors, prior to further analysis, datasets are centered and normalized. As there are many missing values in the dataset, before normalization, missing values are masked. Using these techniques, the convergence time was reduced dramatically, and classification accuracy is improved especially when using SVM.

3.3 Experimental Methods

The outcomes of weighted K-Nearest Neighbors (KNN) [Zha12], Singular Value Decomposition (SVD) based method [GR71], Soft Impute [MHT10], Matrix Factorization [PT94], and Mean average, combined with three state of the art classification algorithms; Support Vector Machines [SV99], Random Forest [LW$^+$02], and Gradient Boosting [OPSS11, LO17, CG16]. To analyze the methods precisely, the experiments have been repeated with varying percentages of missing values. Hyper parameters were carefully adjusted by performing an exhaustive search to adjust each classifier. In this section, methods that have been implemented in this

study have been overviewed. All the methods are implemented using Fancy-impute libraries as detailed in section IV.

The K-Nearest Neighbors imputation method selects subjects with similar feature sets to the subject that has missing values. For example, if a sample S has a missing value in feature Q, this method would select all the subjects which are most similar to S and have the feature Q. It gives a weight to each retrieved sample based on the degree of similarity and then calculates the weighted average as the estimated value for the missing target in sample S. For the similarity measure, various metrics can be utilized such as Pearson correlation, Euclidean distance, and variance minimization. In our study we exploited the Euclidean distance as the similarity measure of the data [Zha12].

Matrix Factorization method was first introduced in [PT94] and since then it has been used in many applications such as collaborative filtering and missing value imputations. This technique attempts to split the original large matrix of $X \in R^{n*m}$ in which n is the number of subjects and m is the number of features, into two matrix components of smaller dimensions as a function of a k factor, $W \in R^{n*k}$ and $H \in R^{k*m}$. Considering that in the imputation problems the original matrix of samples and features has a lot of missing values, the sparsity constraint should be imposed on matrix H which results in the following minimization formulation (3.1):

$$\min_{W,H} \frac{1}{2} \left[\| X - WH^T \|_F^2 + \alpha \| W \|_F^2 + \beta \| H \|_F^2 \right] \tag{3.1}$$

$$\text{subjects } W, H \geq 0$$

with regularizing constants α and β with $\| . \|_F^2$ as the Frobenius Norm. To reach the global minima, the mentioned minimization problem is solved using gradient descent [PT94].

The Singular Value Decomposition (SVD) method has been proposed in [SV99], which is another approach for estimating the missing data iteratively. Assume that X is a set of observed elements and subset of X. SVD-impute applies singular value decomposition of matrix X to get orthonormal patterns of U and V. The approximation of X^r can be derived by a linear combination of these patterns through $J_J D_J V_J^r$ where J_J, D_J and V_J^r are orthogonal. Then, the SVD imputation of any matrix X can be implied by solving the following problem (3.2):

$$Min \parallel X - m_i^r - U_j D_j V_j^r \parallel \tag{3.2}$$

where m_i^r is the mean of the i^{th} row and $\parallel . \parallel$ is a sum of squared values of all non-missing elements. In this method, we start the procedure by substituting the missing values in X by the means of all non-missing values in each row. Then (3.2) will be solved for a new set of matrix of U, V and D which produces a new approximation of X. This step will be repeated until the difference between the X_{i+1} and X_i meets the optimal stopping criteria [GR71].

Soft Impute has been proposed in [MHT10] as a more efficient algorithm than the original iterative SVD which addresses the high computational cost of iterative SVD for large matrices. However, it computes a low-rank SVD of a dense matrix repetitively. This allows the regularization path of solutions to be computed efficiently on a grid of regularization parameters. Rank reduction and shrinkage is performed simultaneously in soft impute in a single operation. More precisely, this algorithm solves equation (3.3) to deduce and replace the missing values. Then the SVD imputation of any matrix X can be implied by solving this problem:

$$\underset{z}{^{minimize}} \frac{1}{2} \parallel W - Z \parallel_F^2 + \lambda \parallel Z \parallel \tag{3.3}$$

where λ is the regularization parameter. This algorithm initializes the missing values with zero and keeps track of the old Z and replaces the Z^{new} with $S_{\lambda_k}(P_\zeta(X)+$

$P^{\frac{1}{\zeta}}(Z^{old})$), until it hits the exit or stop criteria define below (3.4):

$$\frac{\parallel Z^{new} - Z^{old} \parallel_F^2}{\parallel Z^{old} \parallel_F^2} < \epsilon \tag{3.4}$$

$$P_\Omega(X)(i,j) = \left\{ \begin{array}{ll} X(i,j), & if \;\; i,j \in \Omega \\ 0, & if \;\; i,j \notin \Omega \end{array} \right\} \tag{3.5}$$

$P_\Omega(X)$ (with dimension $M * N$) is a projection of matrix X onto the observed entries.$P_{\frac{1}{\Omega}}$ is a complementary projection $P_\Omega(X) + P_{\frac{1}{\Omega}} = X$. The above low-rank optimization models are usually used for collaborative filtering, nonetheless they have application in other domains such as missing data imputation, clustering and data retrieval.

Support Vector Machines (SVM) is a robust statistical method first introduced in the early 1990s as a nonlinear solution for regression and classification [SV99]. This technique has been proven to have superior performance in addressing various problems due to its generalization abilities, robustness against noise and other forms of interference, and its computational efficiency as compared to several other methods. Support vector machines separate two or more classes by finding an optimal hyperplane with a maximized margin known as support vectors. Multi-class SVM problems can be solved by decomposition into a predefined number of binary problems. Two known approaches are one-versus-rest and one-versus-one. One-versus-rest classifiers are composed of k separate binary classifiers in which each classifier will be trained using the data of its own class with a positive outcome and the data from all other classes as negative outcome. One-versus-one approach is composed of all pairwise individual classifiers where each test example will be fed into all individual classifiers and the data will be assigned to the class which yields the highest winning score [OPSS11].

Random Forest (RF) is a type of supervised machine learning algorithm which is an ensemble of multiple decision trees. For each tree in the forest a bootstrap sample of data is taken to create various input datasets so that each tree will be fit with a different set of samples. Then the data will be split based on a selection of random variables. The best split will iteratively be selected based on the impurity measure. The whole process will be repeated in building several decision trees to complete the random forest model. Each new data point will be fed iteratively into all generated trees and their outcome will be averaged to form the final prediction of the random forest [LW+02].

Gradient Boosting (GB) is a powerful supervised machine learning technique based on regression, classification and ranking. This technique has a sequence of weak tree learners which are trained to fit a given model F such that each learner will improve the prediction accuracy of the previous one by minimizing the multiclass logistic likelihood J between the pseudo residuals defined with the following formula (3.6):

$$J = \sum_{i} L(y_i, K(x_i)) \tag{3.6}$$

In which y_i is the target value and $F(x_i)$ is the value obtained from the predicted model. GB is highly robust to redundant data, and has the inherent ability to handle missing data. Hence, we was interested in using Gradient Boosting in our study to examine and test this embedded imputation strength of this algorithm against the proposed cascaded imputation classification method [OPSS11, LO17, CG16].

The experiments proceed in multiple steps, the first is to estimate the missing data using different imputation techniques including KNN impute, iterative SVD, Matrix Factorization, Soft Impute and mean averaging. The next step is the classification of subjects using data that has been acquired from an earlier imputation

step. We randomly selected 70% of the data for training, 10% for validation set, and the rest for testing purposes. We also normalized the data by subtracting the mean value and dividing by the standard deviation prior to imputation. A mask was generating to cover the Not Available (NA) or missing values. We implemented the code in Python using Scikit-learn module for machine learning [PVG+11] and Fancy-impute libraries. While other classifiers were more robust when nonnormalized data were used, SVM accuracy improved dramatically due to normalization.

The diagnosis labels (AD, CN, EMCI, LMCI) were excluded, and some of the highly correlated cognitive test scores such as Alzheimer's Disease Assessment Scale-Cog (ADAS-Cog), Mini Mental State Examination (MMSE), Clinical Dementia Rating scale (CDR) were removed from training. This is worth mentioning that classification accuracy based on the features used for training, especially those used at baseline like MMSE and CDR; and thus, comparative assessments to other studies will be fair only if similar features/modalities and similar datasets are used. In this study, imputation has been done across training, validation, and testing data separately prior to classification. Each classifier has been adjusted through exhaustive grid search with cross validation to achieve optimal accuracy. Tuning parameters for SVM are (Gaussian) radial basis function (RBF) kernel with Gamma and C equal to 0.0001 and 100, respectively. For RF our maximum number of features at each node was set to 10, the minimum number of samples required in each leaf set to 3, and the minimum number of samples required to split an internal node set to 2, with a Gini index for criteria of quality split. For GB we picked out maximum depth of individual regression estimators as 2, the number of features as 25, sub-sample used for fitting learner as 14, minimum number of samples as 10, and number of boosting stages as 28.

To attain a robust performance prediction all experiments have been over 30 tri-

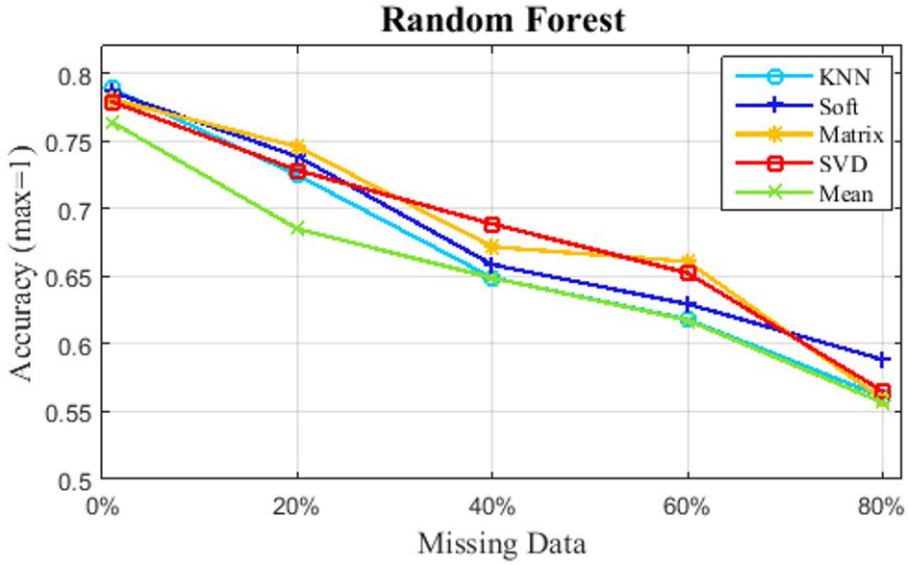

Figure 3.1: Effects of imputation methods coupled with Random Forest classifiers at different degrees of random missing values on 4-way (i.e. multiclass) classification (CN, EMCI, LMCI, AD) of the subjects

als and the metrics across all trials have been averaged. Besides providing accuracy, we also provide performance evaluation metrics that include precision, recall, and Receiver Operating Characteristic curve (ROC) which is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various thresholds. In the following part the experiment is explained in more details.

3.3.1 Random Missing Data Handling

To have a clear understanding of the effect of imputation techniques on diverse patterns of missing values, performance of each technique is evaluated on a complete extracted version of the original dataset where only observations with no missing values were retained. Subsequently, we randomly deleted 1%, 25%, 50% and 75% of the data to investigate the best combination of classification-imputation pairs that

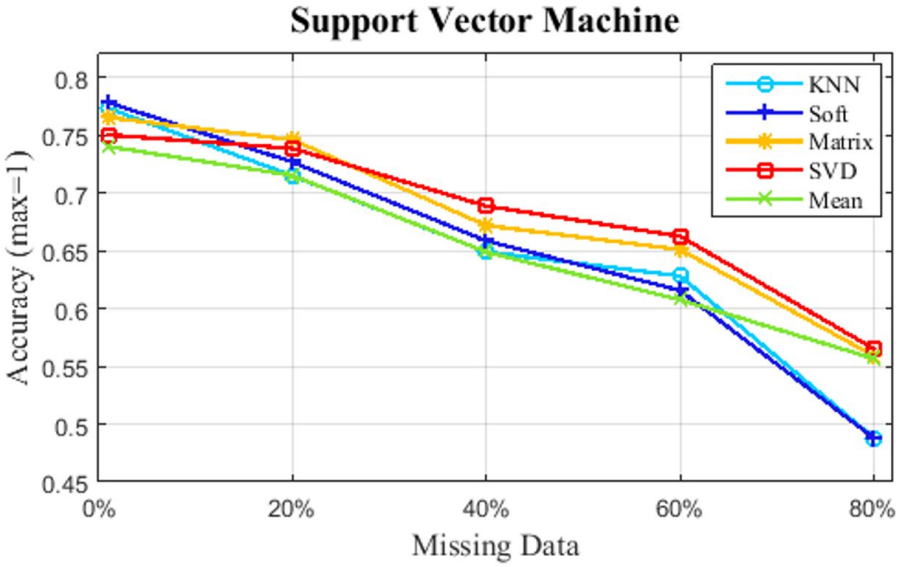

Figure 3.2: Effects of imputation methods coupled with Support Vector Machine classifiers at different degrees of random missing values on 4-way (i.e. multiclass) classification (CN, EMCI, LMCI, AD) of the subjects

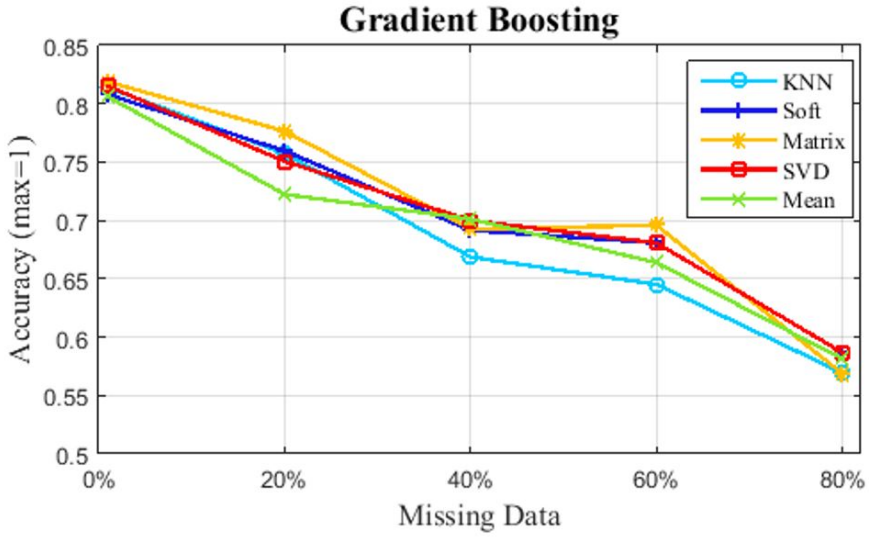

Figure 3.3: Effects of imputation methods coupled with Gradient Boosting classifiers at different degrees of random missing values on 4-way (i.e. multiclass) classification (CN, EMCI, LMCI, AD) of the subjects

achieves optimal classification accuracy. This experiment proves that the SVM and RF classifiers, when coupled with an imputation method like SVD, soft impute or KNN produces the highest accuracy in multiclass classification as compared to mean substitution in almost all percentages of missing values at random less than 80%. The comparison of the results, which is illustrated in figure 3.1 and 3.2 provides enough evidence that selecting the appropriate imputation technique can improve the accuracy of SVM and RF in case of missing data with a random pattern rather than the block pattern. Nonetheless, GB achieves relatively equivalent result with or without the imputation techniques owing to its innate ability for handling missing data. Figure 3.3 also demonstrates that from low to middle percentage of random missing values GB exhibits the highest classification accuracy, regardless of the coupled imputation techniques. To illustrate the effectiveness of different imputation technique in presence of various amount of synthesized missing data, we calculated the accuracy improvement for each classifier as shown in figure 3.4,3.5 and 3.6 for a low percentage of synthesized random missing values, the highest improvement in accuracy is achieved by RF coupled with Matrix Factorization technique as shown in figure 3.4 which happened at 20% of missing data. Additionally, this experiment shows that for a high percentage of missing values, none of the imputation techniques can estimate patterns of missing data correctly.

3.3.2 Original Missing Data Handling

We repeated our experiments on the original incomplete dataset, where almost 40% of its data is missing and the pattern considered in this case is not random but is assumed block-wise [LJ12]. Considering the measurements summarized in table 3.2-3.6, it can be observed that GB method yields the best results over all combinations

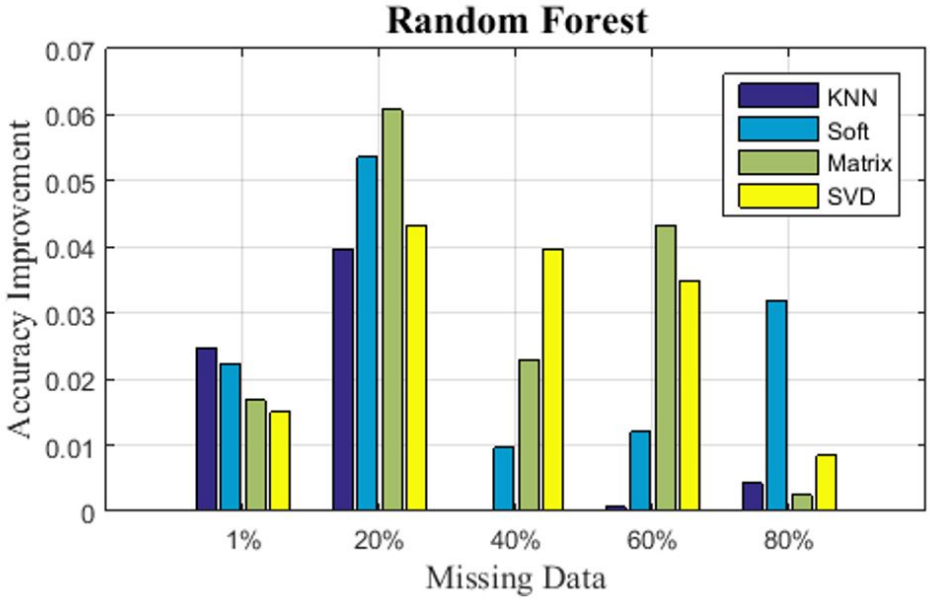

Figure 3.4: Performance of Random Forest on real missing data (blockwise missing) coupled with different imputation techniques

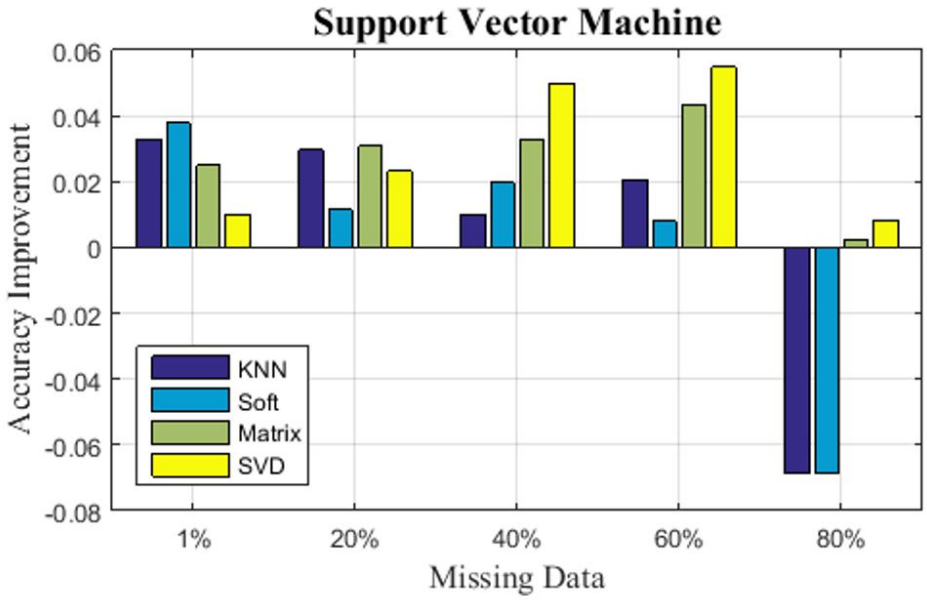

Figure 3.5: Performance of Support Vector Machine on real missing data (blockwise missing) coupled with different imputation techniques

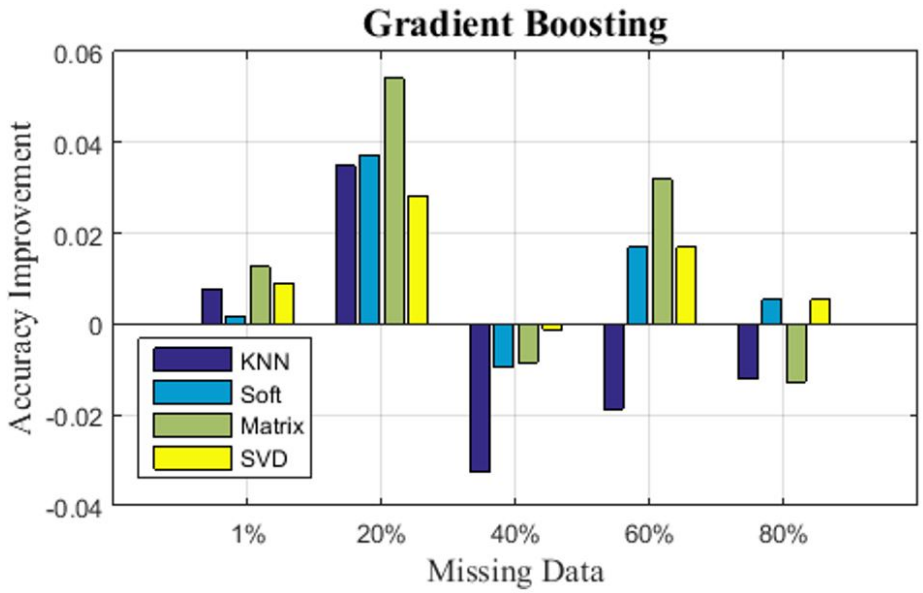

Figure 3.6: Performance of Gradient Boosting on real missing data (blockwise missing) coupled with different imputation techniques

of classifiers and type of value substitution for GB can be observed through the ROC curves in figure 3.7-3.9 The ROC curves of the classifiers represent the difficulty in delineating the four classes (CN, EMCI, LMCI, AD) in a real comprehensive dataset. Among all, GB archives the highest AUC for all the classes which is AUC =0.89, 0.86, 0.84 and 0.78 for AD, CN, EMCI and LMCI respectively while RF with AUC = 0.87, 0.84, 0.82, 0.74 has the second place and SVM with 0.84, 0.82, 0.80, 0.74 shows a somewhat lowest performance for all the classes. figure 3.7-3.9 reveals that EMCI and LMCI separation is the most difficult task for all the three classifiers.

Based on the missing data patterns and also quantity of the missing data, imputation-classification pairing can perform better than simple mean value substitution or no substitution but this improvement highly depends on the distribution of the data and its missing values and cannot be guaranteed to happen in all the cases. Hence, these investigations reveal that none of the state of the art imputation

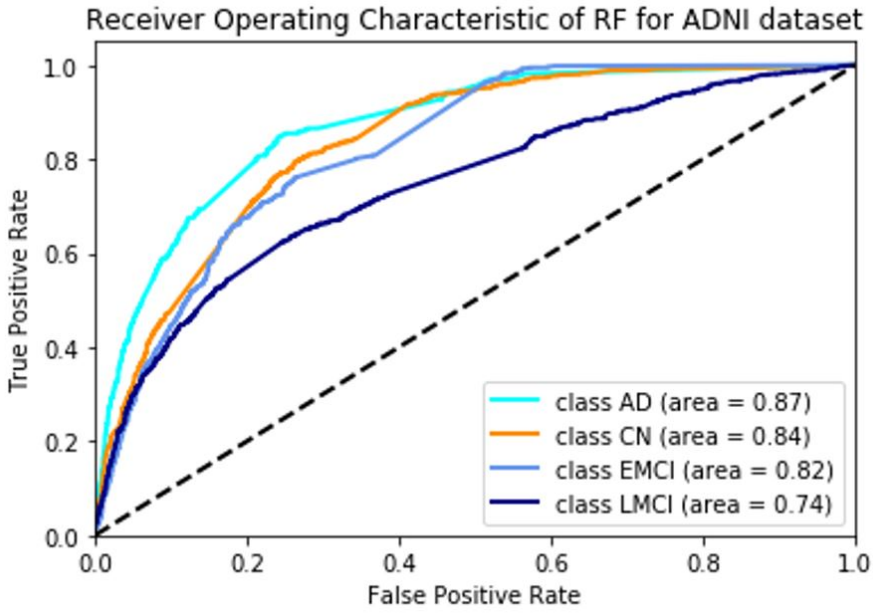

Figure 3.7: Receiver Operating Characteristic of Random Forest for ADNI dataset

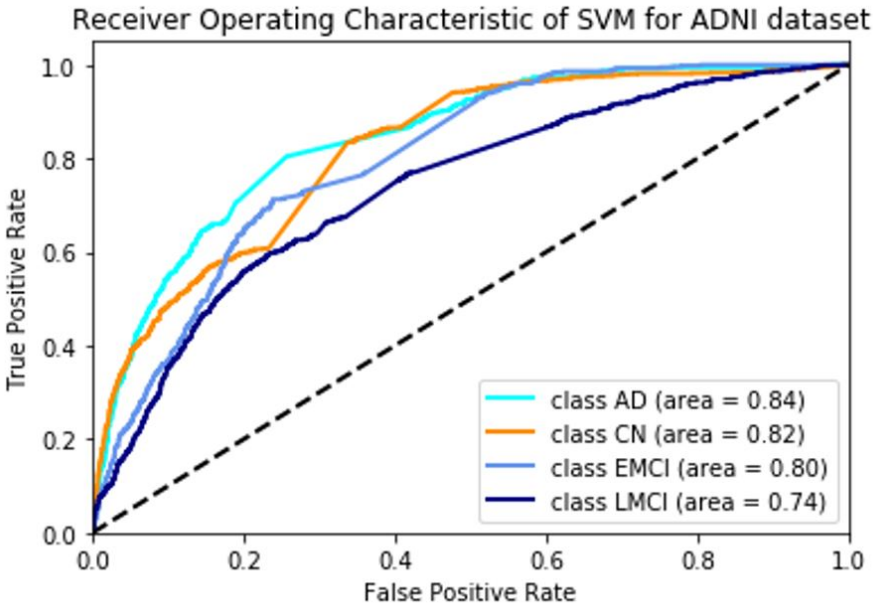

Figure 3.8: Receiver Operating Characteristic of Support Vector Machine for ADNI dataset

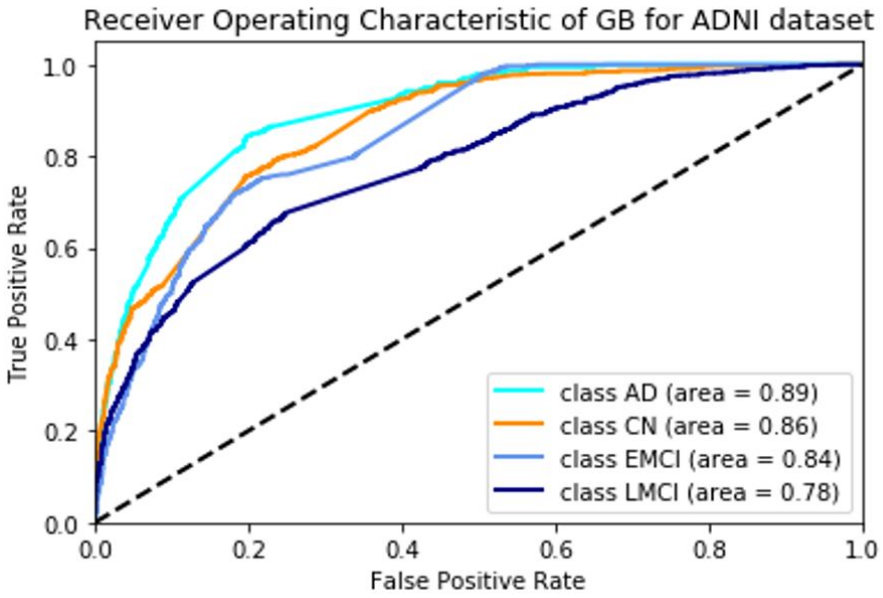

Figure 3.9: Receiver Operating Characteristic of Gradient Boosting for ADNI dataset

techniques have the ability to address block-wise missing data.

To emphasize the difficulty of multiclass classification, the accuracy of binary classification, which has been a main focus of AD related research for many years [XYF+13, GAH+13b], is provided in 3.2 - 3.6. These results highlight that even though classification of subjects between two classes at a time provides higher accuracy, F-score, precision and recall in almost all the cases (AD vs CN, CN vs EMCI, EMCI vs LMCI, and LMCI vs AD) these types of classification lacks as expected the generalization ability for real-world scenarios when an unseen sample data may belong to any of the four groups of (AD, LMCI, EMCI, CN). Four-way classification is more desirable and much more challenging especially when dealing with heterogeneous datasets.

In our study, although GB performed only slightly better than other methods (2% higher accuracy), it holds perhaps the greatest promise because of its versatility,

allowing it to assume simpler, faster and more interpretable forms, such as component wise boosting and the ability to incorporate automatic predictor selection. This study also provides evidence that imputation cost in terms of computational overhead is more rational when the percentage of missing values is under 30% and the pattern of missing data is assumed random.

All algorithms evaluated in this study are robust and successful when considering large feature sets. However, SVM works well for smaller number of observations. RF, on the other hand, is preferable for large non-normalized datasets. SVD and KNN use the correlation structure of the data and KNN uses the Euclidean distance to measure similarity and profile most related observations to estimate the missing values based on this similarity. These approaches will fail to find the most similar profile when it comes to outliers. This flaw can be overcome with scaling or using log over observations. In addition, although the superiority of SVMs against other machine learning algorithms in terms of accuracy has been reported in many studies, this study shows that GB can achieve higher performance in ADNI dataset with its inherent capability of managing the missing values. RF and GB are also quite robust with respect to collinearity. However, SVM alleviates the multi collinearity problem via regularization, where in RF, it is alleviated via choosing a random subset of features for each tree.

3.4 Retrospective

In this chapter, we presented a comparative study of several methods for the estimation of missing values in the largest heterogeneous dataset pulled from various longitudinal studies and cites. We discussed the difficulty of classification in the inherent presence of missing values in longitudinal and multimodal studies and when dealing

with dataset heterogeneity. Of the different state-of-the-art algorithms implemented in this study, Gradient Boosting algorithm achieved the highest performance when dealing with multiclass classification involving all 4 groups (CN, EMCI, LMCI and AD). GB has outperformed SVM and Random Forest algorithms. All the classifiers have been coupled with four advanced imputation techniques including KNN impute, Matrix Factorization, SVD, and Soft Impute and they have been utilized to classify the different stages of AD. Despite the contribution of the imputation techniques in missing value estimation in data with low percentage of the random missing data, all the algorithms fail to perform well in high levels of missing data. Moreover, in the presence of block-wise missing data patterns, where a particular modality is completely missing for so many subjects, these imputation methods are not helpful. While many studies so far focused on binary classification of AD, we went further in performing multiclass classification while contending with the missing data challenge inherent to longitudinal and multimodal studies.

Moreover, We also provided results of the different binary classifications as well for comparative purposes and for estimating the effect of missing data on such binary classification in contrast to multiclass classification. The imbalanced dataset and insufficient samples in each group of subjects imposed a new constraint on the current classification problem. We tried to tackle this issue by incorporating the data samples from longitudinal studies and provided effective ways to augment the dataset.

Table 3.2: Comparison Of (A) Random Forest (B) Support Vector Machine (C) Gradient Boosting coupled with five imputation techniques in 4 way classification task

Classifier	Gradient Boosting			Support Vector Machine			Random Forest		
Imputation tech	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
KNN	59.21	60.5	57.1	58.89	51.8	55.29	53.65	56.82	51.61
Soft Impute	65.03	63.12	63	58.88	54.61	57.91	57.54	58.1	55.65
Matrix Factorization	62.1	60.65	59	57.48	48.92	56.34	55.61	57.02	53.1
Iterative SVD	62.2	62.27	62.2	58.83	60.49	56.88	58.18	55.91	56.41
Mean	**62.12**	**63.32**	**62.2**	57.7	52.58	55.47	60.56	60.59	59.65

Table 3.3: Binary classification of the Control Normal vs Early Mild Cognitive Impairment (CN vs EMCI)*

Classifier	Gradient Boosting			Support Vector Machine			Random Forest		
Imputation tech	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
KNN	70.28	70.54	67.2	78.18	77.39	77.01	72.01	70.14	70.11
Soft Impute	79.07	76.79	76.27	76.91	75.19	75.15	73.71	72.39	72.47
Matrix Factorization	80.23	77.21	77.21	76.54	73.45	73.44	75.85	73.54	73.28
SVD	80.19	79	79.51	76.08	74.54	72.88	74.97	73.72	73.94
Mean	**80.93**	**79.67**	**79.9**	76.96	75.93	76.22	75.56	73.91	74.33

Table 3.4: Binary classification the Early vs Late Mild Cognitive Impairment (EMCI vs LMCI)*

Classifier	Gradient Boosting			Support Vector Machine			Random Forest		
Imputation tech	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
KNN	83.19	81.32	81.98	83.13	79.11	80.34	79.98	78.45	79.28
Soft Impute	82.75	79.53	80.71	82.42	81.72	82.34	83.56	80.67	81.63
Matrix Factorization	85.12	81.14	82.16	77.35	76.66	77.44	81.55	79.73	79.54
SVD	85.81	81.12	82.33	79.12	77.65	79.29	81.61	80.13	80.26
Mean	**85.84**	**82.22**	**82.4**	81.38	81.2	82.4	83.24	80.74	81.74

Table 3.5: Binary classification the Late Mild Cognitive Impairment vs Alzheimer Disease (LMCI vs AD)*

Classifier	Gradient Boosting			Support Vector Machine			Random Forest		
Imputation tech	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
KNN	73.05	67.24	69.53	65.55	46.44	64.07	73.92	70.78	72.2
Soft Impute	74.84	71.08	72.42	69.92	47.79	68.4	73.45	69.98	71.4
Matrix Factorization	73.84	70.87	71.81	66.25	43.58	63.97	72.32	68.31	70.2
SVD	74.23	70.42	71.34	69.32	46.75	69.47	74.63	72.46	72.6
Mean	**75.23**	**73.16**	**73.35**	70.12	44.35	66.87	74.59	71.67	72.29

Table 3.6: Binary classification the Control Normal vs Alzheimer Disease (CN vs AD)

Classifier	Gradient Boosting			Support Vector Machine			Random Forest		
Imputation tech	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
KNN	93.3	91.54	91.14	91.82	90.78	90.24	92.77	90.52	90.14
Soft Impute	91.67	91.36	91.07	90.89	91.02	91.08	91.28	89.61	89.83
Matrix Factorization	91.41	91.32	91.34	91.84	90.84	90.18	90.74	90.74	90.85
SVD	90.9	90.88	90.29	90.57	90.61	90.75	89.81	89.15	89.18
Mean	**93.4**	**92.37**	**91.44**	91.85	90.11	89.54	92.71	90.21	90.12

CHAPTER 4

PREDICTIVE MODELING OF LONGITUDINAL DATA FOR ALZHEIMER'S DISEASE DIAGNOSIS USING RNNS

4.1 Introduction

Regardless of enormous efforts, pinpointing the prodromal stage of mild cognitive impairment is remained an open research field. Having incomplete samples in the longitudinal medical studies is a common phenomenon, as many patients may miss some of the tests and modalities in a time step or miss a complete visit within the study's lifespan. Generally, missing values occur for a variety of reasons including drop out of subjects from the study, insufficient resolution, image corruption, budget limitation, etc. [MPG+15b, NZM+16, ZYLY11b]. Many algorithms simply discard subjects with missing modalities from further experiments,which indeed results in a considerable loss of valuable information. Disease diagnosis accuracy might be improved if the missing parameters could be estimated correctly from the rest of the available data or modalities. Furthermore, to have a better understanding of the disease progression and to correctly label a subject as Normal Control (NC), Mild Cognitive Impairment (MCI),or dementia (i.e.,AD), data from every visit should not be scrutinized independently from the earlier steps. Currently,a majority of the classification algorithms focus on the cross-sectional data and only analyze a specific interval's biomarkers for the diagnosis and disregard the former patient's status for the decision making process. To address this shortcoming, recent studies moved toward longitudinal data analysis and proposed new methods to leverage valuable temporal data by considering the inherent correlations of such data [NZM+16, ZYLY11b, ZSI+12b]. Effectively mining AD longitudinal data is a challenging task, owing to its heterogeneous measurements, varying length of sam-

Table 4.1: Dataset statistics

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
NC	521 (268/253)	74.25±5.79	16.37±2.70	29.06±1.14

ples, missing modalities and tests,and small sample size. In this study, for the first time(to the best of our knowledge), two RNN models, namely the Long Short Term Memory (LSTM) and the Gated Recurrent Unit (GRU) are employed to discover the regression patterns of the subjects from the longitudinal data with missing variables and intervals,especially for the task of classifying AD/MCI vs.NC,which is a challenging task only depending on the cross-sectional dataset. The progression of the patients during time should be studied care-fully to capture the correct status of the patient through the passage of time. Accordingly, in this study, Several experiments have been conducted to investigate the effectiveness of the RNNs in AD diagnosis. The outcomes of the LSTM and GRU model with Multi-Layer Perceptron (MLP) were compared to evaluate the efficacy of the sequential models.

4.2 Dataset

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. Recently the largest longitudinal dataset,which

is a subset of ADNI 1/Go/2 cohorts, has been extracted from ADNI by Bruno M.Jedynak and Michael Donohue to make a baseline for researchers in the field to propose and apply quantitative templates for the progression of Alzheimer's disease. This is an invaluable baseline for accurate evaluation of the proposed algorithms. The database has 1721 distinct subjects (521 NC, 864 MCI, and 336 AD)examined every 6 months during 11years' period making 23time points for a patient in the case of performing all the test regularly every six month(i.e., baseline,6 months, 12 months,..., 132 months). For every visit multiple outcomes provided including ADAS13, CDRSB, RAVLT.learning MMSE, FAQ, FDG PET, Amyloid PET, CSF, ABETA, CSF TAU, CSF PTAU, FS WholeBrain, FS Hippocampus, FS Entorhinal, FS Ventricles, FS MidTemp, FS Fusi form and the covariates: age, APOE4 (yes/no), Gender, Education. The primary phenotype is the diagnostic group and Mini-Mental State Examination (MMSE). Sample data-point curation pipeline in our work is presented in figure 4.1. This figure shows that the samples are composed of features extracted from volumetric magnetic resonance imaging (MRI) including cortical thickness, hippocampal volume and shape along withfluoro-2-Deoxy-D-glucose, florbetapir F18, and PIB (which is radiotracer capable of highlighting deposits of beta-amyloid) from PET imaging,and some other Cerebrospinal fluid(CSF) features such as TAU, PTAU and ABETA. Around 12 functional and behavioral assessment results such as Rey's Auditory Verbal Learning Testand Montreal Cognitive Assessment (MoCA) scores are also measured and used as features in this dataset. The volumetric MRI measurements provide the cortical thickness, volume and shape of hippocampal or voxel-wise tissue probability [GW84, HDD$^+$94, CGT$^+$11b, PM05b] to measure the brain atrophy; 18-Fluoro-DeoxyGlucose PET imaging (FDG-PET) estimates the glucose hypo metabolism in bilateral temporal, temporal occipital areas or posterior cingulated brain regions [MPG$^+$15b, NZM$^+$16, ZYLY11b]. Fur-

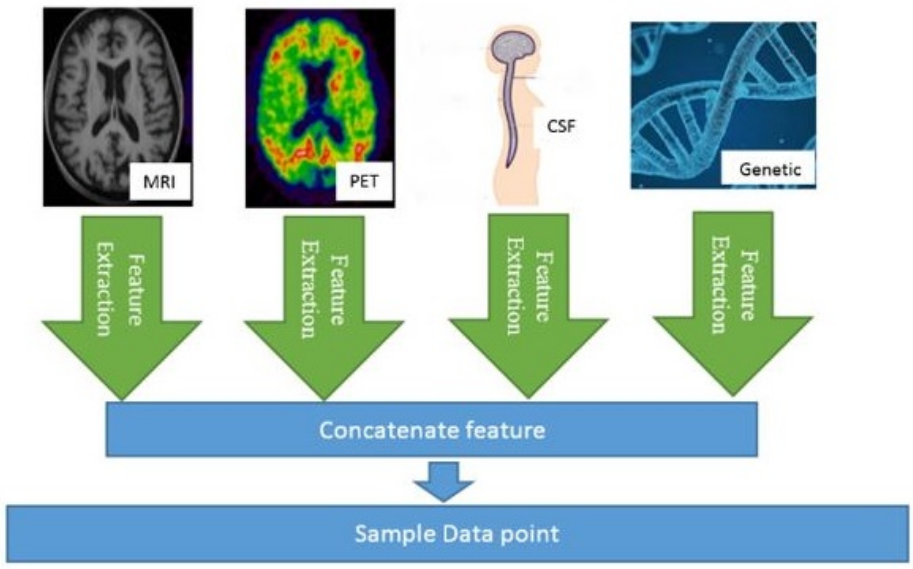

Figure 4.1: Sample datapoint curation

thermore, global cognitive impairment tests are used by clinicians for screening and measuring individuals who are at the risk of AD; or cerebrospinal fluid(CSF) to measure the increase in t-tau, p-tau, or the decrease of amyloid-beta, which is a sign of cognitive declination. Therefore, in total 47 features are used to represent each subject at each time point. Dataset statistics has been provided in table 4.1

4.3 Models

In this section, we briefly overview the LSTM and GRU models used in our model and then explain our model design using these architectures for classifying the subjects into one of the AD, MCI, or NC categories from longitudinal data.

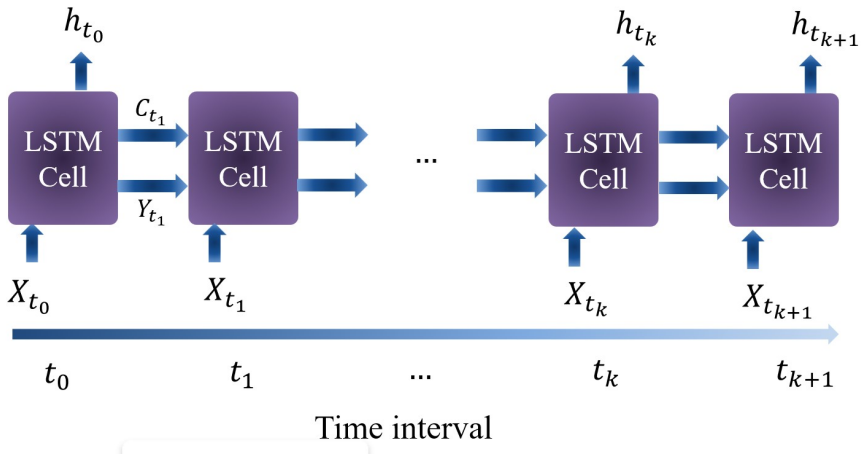

Figure 4.2: An RNN network with LSTM cell with (k+1) time points.

4.3.1 Long Short Term Memory Unit (LSTM)

RNNs with internal memory and feedback loop have previously been adopted mostly for processing arbitrary input sequences, like in hand writing recognition, speech recognition, natural language processing, and time series prediction applications. One of the main challenges in applying RNNs to long sequential data is that the gradient of some learnable weights become too small or too large if the network is unfolded for too many time steps. These phenomena are called the exploding and vanishing gradients problem [BSF94]. LSTM was, hence, proposed by Hochreiter et al. for the first time in 1997 to solve the vanishing gradient problem through a gating mechanism [HS97]. An LSTM has three gates. The first gate determines whether the information should be for gotten or not. The second gate decides about updating the cell state,and the last gate is responsible for the cell output. Since then,several variations of LSTM architecture have been implemented especially with the utilization of Graphics Processing Units (GPUs). An RNN network with LSTM cell with (k+1) time points has been depicted in figure 4.2

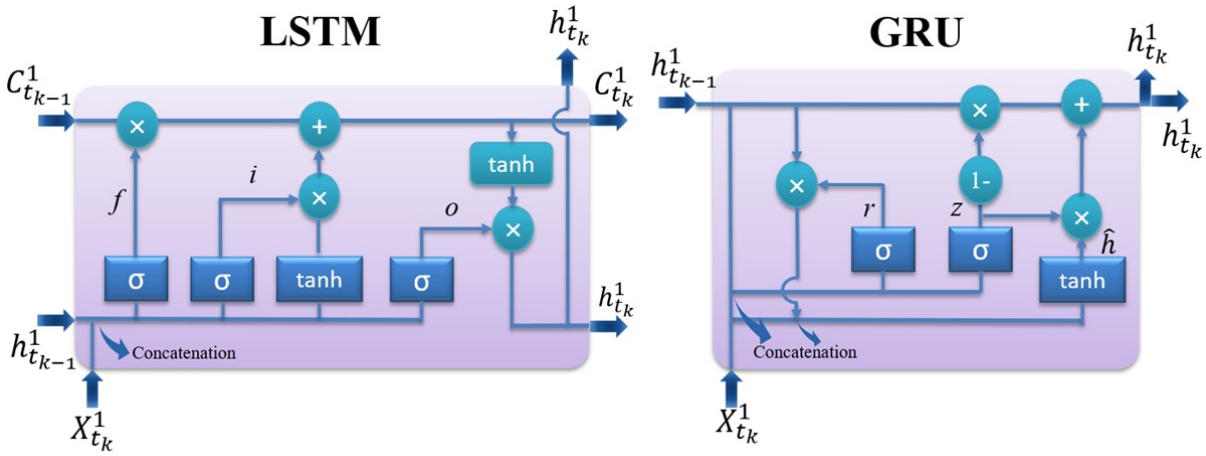

Figure 4.3: The structure of LSTM and GRU cells

4.3.2 Gated Recurrent Unit (GRU)

To adaptively capture dependencies of different time scales in each recurrent unit, Cho et al. [CVMG+14b] introduced a gated recurrent unit (GRU). Similar but not the same as LSTM design, GRU has two gates, a reset gate r, and an update gate z. Intuitively, the reset gate determines how to combine the new input with the previous memory, and the update gate defines how much of the past memory to keep around. Having simpler architecture than LSTM with a smaller number of parameters, GRU provides better results in some applications [CPC+18] and is less prone to over-fitting, especially in cases that there are not enough training data. The structure of LSTM and GRU have been shown in figure 4.3

4.3.3 Proposed Model

RNN models have achieved popularity due to their power in pattern recognition for the time series and sequential data. While there are plenty of research papers on regression and classification modeling of AD data with well-established and novel

machine learning techniques, along with many deep convolutional neural networks for 2D and 3D brain MRI classification, number of research works exploiting RNNs for finding the patterns in the AD longitudinal data sets is limited [SCZ+17, FLC+17, SZL+17, CRG+09, ZSW+17, LWVW+14]. Only a few papers recently adopted them for regression analysis on the clinical medical data[BEWY]. Here, a RNN deep learning technique is employed for the classification of the subjects. All features are normalized by subtracting the mean value of each feature and dividing the result by the standard deviation of that feature in all samples (i.e., using their z-scores),before the analysis. To deal with missing modalities, we replace them with zero values. Since our goal is to showcase the usage of RNNs for longitudinal predictive analysis, we leave extensive data imputation experiments for future works. A recent work also models AD progression with RNN models[CLL18]; however our work is different from that in multiple aspects. We not only use MRI features but also PET, Cognitive tests, and genetic features for modeling the disease. We also propose multiple approaches for handling the missing intervals and compare the potential RNN models with each other. The dataset contains N=1721 subjects each scanned in 24 different time points. Data from each time point is represented by n = 47 features. Figure 4.4 overviews the data arrangement. A challenge in analyzing longitudinal data sets is dealing with missing data at different time steps for some of the subjects. To address this inconsistency in the data points and to be able to input the data to RNNs,we define three settings:(1) In our first attempt, we fill the missing intervals with zero to create a same input size data for all the subjects and compose a stack of 1721, 2D matrices that all have a set of 47 biomarkers in the columns as features and all the possible time steps in the rows as time steps. We refer to this arrangement as zero fill. (2) In the second attempt, we buffer the data at every time point and replicate it in its next missing inter-val. This scheme is

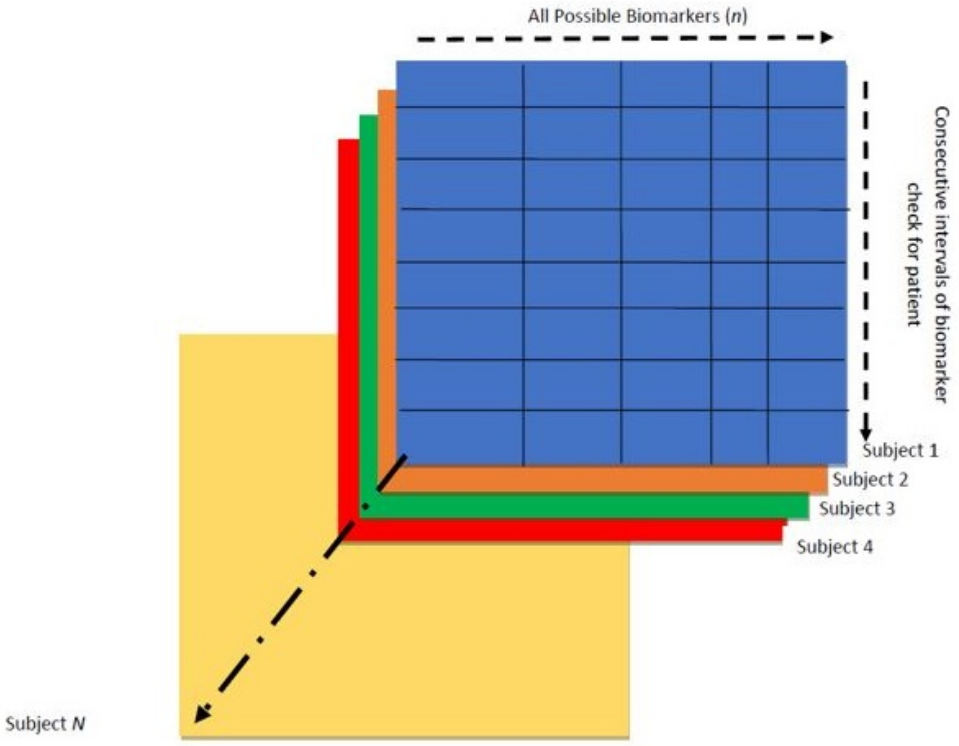

Figure 4.4: Data arrangement for RNN model

named as replicate fill. (3) In the last configuration we change the orientation of the input data and stack all the available intervals on top of each other, disregarding the missing intervals and pad them to the maximum size of the possible time steps, this is called padding.One LSTM and GRU model with the memory of the maximum size of the available time steps, which is 24, are designed to process this stack of data. Each subject's time point data is fed to the corresponding cell along with its final diagnosis label (i.e.,AD,MCI, or NC) allowing the model to learn the pattern of the change in the features for each subject. Figure 4.5 represents this pipeline. In two different sets of experiments, we replace the cells in this figure with LSTM and GRU sets and report the results.

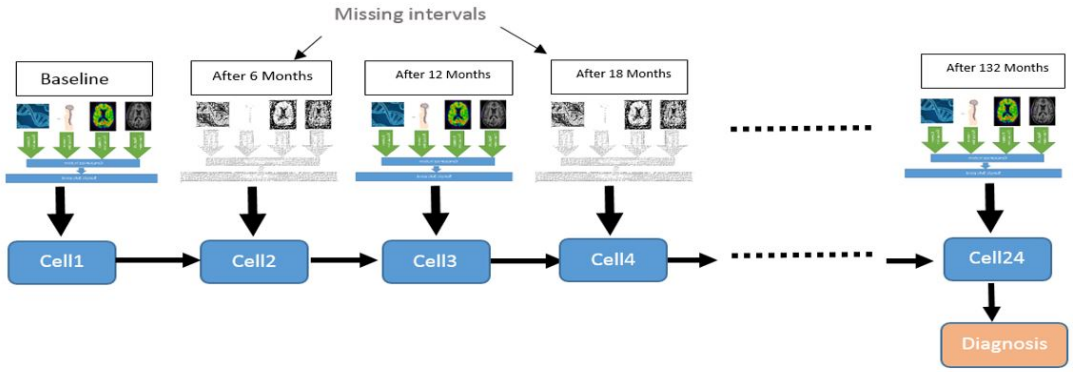

Figure 4.5: Schematic of RNN model and how input data feed to the model

4.4 Experiments

In all the experiments, we train and tune the RNN model with different configurations of the hidden layers, percentages of drop out, various activation functions, loss, optimizers and different combination of other hyperparameters to find the best setting of the model through a grid search. We knowingly made the models as small as possible to avoid overfitting,which can easily mislead the comparison. Data has been split into 70% training, 15% validation set and the rest for the testing set. The best configuration of the LSTM and GRU is represented in table 4.2. For evaluations, we calculate the Accuracy, Sensitivity, Specificity,and F-score of all models. The results of LSTM and GRU models for all arrangements of the data are compared in table 5.4-4.5,along with the results of their counterpart from non-recurrent networks, i.e.,Multi-Layer Perceptron (MLP). The data is flattened to a 1D long vector and fed into the MLP once for each patient. According to table 5.4-4.5, LSTM and GRU models are superior to the MLP network in most of the cases as they result in the highest accuracy and F-score.Our LSTM model yields nearly 1% accuracy improvements over MLP in classifying AD patients from NC subjects. Interestingly,

Table 4.2: Model hyperparameters

	Hidden Units	Activation Function	Layers	Drop out
GRU	32	Softmax	1	0.3
LSTM	30	Softmax	1	0.4
MLP	20	Softmax	2	0.3

the RNN models with the zero fill data arrangement for the missing data yields consistently better results. The superiority is not significant, which can be mainly due to the limited amount of data in this domain,besides the high portion of the missing time points and modalities. These challenges prevented the vanilla RNNs to find the appropriate patterns despite various input data arrangement. Second, RNNs,especially the LSTM models,have a large number of trainable parameters, which necessitate the model to be trained in a great corpse of sequential data and despite having dropout layers in the architecture,they are still prone to overfitting to the training data in this relatively small dataset. The third is the limited hand engineered and structured feature set,used in this experiment. One of the main superiority of the RNNs is their power in automatic feature learning from the raw data,which can be further explored in the future.

Table 4.3: Performance of the proposed models with three different data arrangements in classification of ADNI subjects between two classes of AD vs NC. Best results for each data arrangement are underlined, and the best overall results of each column are in bold.

	Method	Accuracy	F-score	Sensitivity	Specificity
ZERO FILL	MLP	0.9467	0.9581	0.9626	0.9194
	LSTM	0.9526	0.9622	0.9532	0.9516
	GRU	**0.9527**	**0.9630**	**0.9720**	0.9194
REPLICATE FILL	MLP	0.9467	0.9577	0.9533	0.9345
	LSTM	0.9586	0.9674	0.9720	0.9355
	GRU	0.9527	0.9626	0.9626	0.9345
PADDING	MLP	0.9467	0.9577	0.9531	0.9355
	LSTM	**0.9527**	**0.9623**	**0.9533**	**0.9516**
	GRU	0.9408	0.9528	0.9439	0.9355

Table 4.4: Performance of the proposed models with three different data arrangements in classification of ADNI subjects between two classes of AD vs MCI. Best results for each data arrangement are underlined, and the best overall results of each column are in bold.

	Method	Accuracy	F-score	Sensitivity	Specificity
ZERO FILL	MLP	0.8474	0.8449	0.9405	0.7736
	LSTM	0.8579	0.8492	0.9048	0.8208
	GRU	0.8368	0.8360	0.9405	0.7547
REPLICATE FILL	MLP	0.8529	0.8492	0.9048	0.8208
	LSTM	0.8576	0.8498	0.9286	0.8225
	GRU	0.8211	0.8211	0.9286	0.7358
PADDING	MLP	0.8421	0.8295	0.8690	0.8208
	LSTM	0.8468	0.8298	0.8810	0.8219
	GRU	0.8158	0.8108	0.8929	0.7547

Table 4.5: Performance of the proposed models with three different data arrangements in classification of ADNI subjects between two classes of NC vs MCI. Best results for each data arrangement are underlined, and the best overall results of each column are in bold.

	Method	Accuracy	F-score	Sensitivity	Specificity
ZERO FILL	MLP	0.7729	0.7539	0.6207	0.9670
	LSTM	0.7729	0.7793	0.7155	0.8462
	GRU	0.7536	0.7536	0.6724	0.8571
REPLICATE FILL	MLP	0.7005	0.6667	0.5345	0.9121
	LSTM	0.7681	0.7757	0.7155	0.9352
	GRU	0.7101	0.7000	0.6034	0.8462
PADDING	MLP	0.7101	0.7609	0.6877	0.8423
	LSTM	0.7585	0.7619	0.6897	0.8462
	GRU	0.7101	0.7000	0.6034	0.8462

CHAPTER 5

DEEP LEARNING BASED SEGMENTATION MODELS FOR ALZHEIMER DISEASE DIAGNOSIS

5.1 Introduction

Hippocampus atrophy is among the most informative early diagnostic biomarkers of AD. The hippocampus is a region of interest (ROI) for various research studies that include memory function analysis, stress development observation and prediction of neurological and neurodegenerative disorders. Hippocampus atrophy expressed through accurate volume calculations could help in both multiclass classification and prediction. Magnetic Resonance Imaging (MRI) has shown to be a promising non invasive modality for AD diagnosis through in-depth assessment of the brain structure, especially of disease-prone areas, among them the hippocampus. Neurologists often use 3D MRI of the patient's brain to diagnose and monitor diseasse progression, and to formulate patient-specific treatment plans in the early stages of the disease. In order to determine the brain morphological changes, regions of interest (ROIs) should be first segmented in MR images. Currently this process is performed manually which is not only tedious and time consuming but also subjective, error prone and non-repeatable [FSB+02]. However, manual segmentation, when performed by two or more experts, can serve as the gold standard for comparing volumetric results obtained from automated segmentation techniques [JJAK+11, WBP+17, BBG+15, BBM+15] as was the case for this 2019 Segmentation Challenge data. These studies on standards and training labels clearly indicate that although there is strong agreement among expert tracers, the need for automated segmentation methods is highly desirable. This automation process is important not only to overcome the tediousness and the labor intensive nature of manual tracing,

especially in 3D medical images, but to also augment the practicality of volumetric calculations of disease prone regions for early detection, multiclass classification, assessing disease progression and performing prediction on future outcomes. Therefore great attention has been devoted in the literature to automate this 3D segmentation process. It is important to emphasize however that regardless of the ways this task is performed, manually or automatically, the burden of proof lies in the fact that beyond the imbalance problem between ROI and background information (i.e., the ROI appears in a very small portion of the entire dataset), there are no clear boundaries between the different regions of the brain, which makes this segmentation endeavor even more challenging. The complexity of this task is best expressed in [BBG+15], showing that even when there is very high agreement among the four expert tracers (pairwise Jaccard indices 0.82-0.87), the volumetric results obtained on the HarP benchmark dataset containing 135 MRIs still showed a mean volume difference of 9% with a standard deviation 7%.

Currently the most accurate and effective semantic segmentation models are based on fully convolutional networks (FCN)[LSD15]. These models are composed of several convolutional layers followed by pooling layers which gradually expand the receptive field and create high-level semantic features. However, by passing through a pooling layer, the size of the feature map is reduced and small objects or nuance changes in the pixels get lost. This leads to an inaccurate boundary reconstruction in the segmentation map. To address this drawback, multi-scale feature maps have been used in the segmentation networks. This technique helps to effectively aggregate complementary information from various scales and complements the missing boundaries of feature maps[HCH+17]. Among many of these CNN models with the multi-scale feature maps are the DeepLab and PSPNet architectures. UNet-based architectures on the other hand have shown superior performance. UNet-based ar-

chitectures with deep feature representation, contraction-expansion path and skip connections have outperformed other models in medical image segmentation tasks, where fine grained masks and accurate margin identification are highly advantageous. While efficient feature extraction with the UNet-based networks has shown great success in image segmentation tasks such as in kidney tumor segmentation, lung tumor segmentation, etc., automatic hippocampus segmentation remains a challenging task due to the low contrast of the surrounding tissue, complicated further by it's small size and irregular shape. More specifically, the small size of the hippocampus compared to the whole brain creates an imbalance between ROI and background information, compromising as a consequence the performance of any segmentation model. To overcome this imbalance problem, this study presents a two-step hippocampus segmentation framework in which the first step is to locate a candidate region surrounding the hippocampus structure in the brain. The candidate region expected to contain the hippocampus will have relatively equal ROI and similar background voxels, attenuating as a consequence the imbalance data problem and improve the segmentation performance. After extracting the voxels containing the hippocampus, the second step is to leverage an efficient segmentation model based on the combined Transformer and UNet architecture to generate the segmentation mask. This architecture can efficiently model global contexts without losing localization ability as the low level details are maintained. We have adjusted the architecture and fine tuned the model to achieve high performance in the hippocampus segmentation task. The hippocampus volume, which is captured almost in real time from this model in the testing phase, can be utilized to enhance AD diagnosis. More importantly, the proposed segmentation algorithm can be deployed on other datasets of interest in pursuit of similar research goals that rely on segmentation and volumetric measurements.

To evaluate the merits of the proposed segmentation task, we further propose multiple combination of Unet, Unet++, MANet and FPN networks coupled with strong feature extraction backbones pre-trained with large datasets of ImageNet or Instagram. The performance of the aforementioned models on the hippocampus segmentation task are thus provided and drawbacks and strengths of each model are discussed in Section IV. In almost all of the models, the pipeline starts with slice-wise feature extraction of voxels attributed to the hippocampus region to be mapped to the target binary mask of structural MRI. In this method, the slice-wise voxels are considered as the volumetric features. The proposed CNN architecture learns the hidden representation of slice-wise volumetric features and ultimately calculates a segmentation mask for the brain hippocampus region.

The rest of this chapter is organized in the following manner. In next section, related literature in this challenging research field is presented. The methodology along with the dataset, data processing, network architecture, and loss functions used in this architecture are discussed in section III. Experimental results are provided in section IV starting with an introduction to the computational platform used to run these experiments followed by a performance evaluation of the proposed method with other state-of-the art methods. Section V provides a discussion with concluding remarks on the contributions of this research work and current limitations that need to be overcome for future neuroimaging segmentation tasks.

5.2 Related work on Hippocampus Segmentation

This section focuses on how the hippocampus region is located and then segmented. Research findings suggest that hippocampal volumetry is an important quantitative metric that can serve as a biomarker for neurodegenerative diseases like Alzheimer's

disease [LHW+19, MID+19, WNK+21, LMW+21], Parkinson's disease [WDD+12] neurological disorders like epilepsy [HWB+04], neuro-psychological disorders like bipolar disorder and schizophrenia [OJMM19] [SLPG04] and depression [HSB+21] [BNA+00], to name a few. Hence, segmenting correctly this region is of utmost importance for assessing any subtle structural changes and thus volumetry that may be due to the pathology. In fact, volumetric alterations and hippocampal degeneration could be useful in the classification of the different stages of AD [CL18, KYP+18, WWS+07], and in distinguishing between AD patients from dementia with Lewy bodies (DLB) patients [WWS+07], and AD from patients with subcortical ischemic vascular dementia as an other form of dementia [DSL+02].

5.2.1 Locating the Hippocampus in Brain 3D MRI

To estimate the hippocampus location in the brain MRI, several techniques have been proposed. Bender et al. [BKB+18] have proposed a technique that first applies both rigid and contour-based image registration on the brain images (CT or MRI), then generates a population-based hippocampal atlas by mapping the hippocampus from several patients into a template image set. Estimated hippocampal contours can be automatically formed in a given image set by mapping this atlas onto it. Multiple researchers have proposed similar segmentation techniques based on the brain atlas registration [CAD+05, vdLDHBN08, WT21]. Thus, techniques based on brain image registration became the most widely used brain anatomical segmentation techniques. Such techniques are adopted in many robust software packages such as FreeSurfer. FreeSurfer explicitly starts with some pre-processing steps including motion correction, affine transformation to image space, non-uniform intensity normalization, and removal of non-brain tissues. Then image volume will be intensity

normalized to match the brain atlas image intensity histogram. The final step to perform brain structure segmentation is a nonlinear warping of the atlas brain image to the sample brain image. This technique is mostly utilized for atlas-based tissue segmentation, in labeling the brain stem, subcortical structures, cerebellum, and cerebral cortex [Fis12].

An intuitive approach has been proposed by Hajiesmaeili et al. [HA17]. Their algorithm performs 3D skull stripping and extracts the brain volume first followed by a distance estimation from the first slice of the brain to the first slice where the hippocampus appears in all 3 views of coronal, sagittal and axial. This algorithm is particularly beneficial if a rough estimate of the hippocampus structure is needed.

Basher et al.[BCL+19] have presented a two-stage process where the model first locates left and right hippocampal tissues in the MRI with the Hough-CNN model, then slices of the hippocampus are sent to a Discrete Volume Estimator CNN model to extract features from both hippocampal tissues. All the features are then aggregated and passed to a final deep neural network where the AD classification happens. Suk and Shen, on the other hand, propose to combine latent information (more complex information of the low level features) with the original low-level features to help build a robust model for AD/MCI classification with high diagnostic accuracy[SS13].

5.2.2 Hippocampus Segmentation

Although manual segmentation or tracing of the hippocampus region remains the gold standard, there is clearly a need for automating the segmentation process [WBP+17, BBG+15, MID+19, WNK+21]. Automatic hippocampus segmentation techniques are mainly based on two types of approaches: 1) Conventional image

processing techniques and atlas registration [BBL+07, HMC+00]; and 2) CNN-based approaches for feature extraction and volume estimation. While many useful methods and software packages have been proposed based on brain multi atlas segmentation, the performance of such models highly relies on registration accuracy, atlas selection, type of brain atlas used and label fusion. These factors reduce the model applicability for general use cases. With the success of deep learning models in various application domains of artificial intelligence, image processing and computer vision, the biomedical research field shifted it's attention toward a new line of research with the advent of the CNN-based models, some of which have been proposed for Hippocampus segmentation [PPW+14, AHH+09].

To overcome the high dimensionality of the 3D brain images, Liu et al. [LZAS18] introduced a landmark-based feature learning approach. In their model, critical informative anatomical brain regions are first detected with a heuristic algorithm. Instead of feeding the entire 3D image to the model, patches of the image surrounding the landmark are fed to a CNN . Several earlier methods try to investigate whole brain features together. In contrast to many models that are trained on the entire 3D MRI, which impose huge computational burden for training with specific hardware needs, this approach needs far less computational power and is easier to carry out the training phase [LZAS18].

Zach et al. [ZBL+20] have proposed a simplified protocol for hippocampal volume change measurement. They have found a single optimal slice of the brain MRI where the hippocampus is the only visible organ. Then by calculating the area and volume of the hippocampus in this specific slice, they have tried to classify 40 subjects in normal control group against 40 subjects in AD group with these biomarkers. They have shown that this simple process can substitute for the complex methods of hippocampal volume estimation without any further needs for brain or skull normalization.

Yanrong Guo et al. [GGS15] have proposed a segmentation model based on rich features which were extracted by stacked sparse auto-encoder from prostate MR images using a sparse patch matching process in a supervised fashion. They showed that the deep-learned features were more effective than the handcrafted features in guiding MR prostate segmentation. A dense V-network has been introduced by Eli Gibson[GGH+18] for multi organ segmentation of abdominal organs. This model is a fully convolutional neural network with five main features, which include: 1) V-shape down sampling and up sampling path 2) dilated convolutions 3) dense feature stacks 4) batch-wise spatial dropout and 5) explicit spatial prior. With this design, they have been able to surpass the dice coefficient metrics of other models such as VNet and VoXResNet with considerable margin in multi-organ segmentation task [HLVDMW17].

5.2.3 CNN Models for AD Diagnosis

Hippocampus segmentation research for AD diagnosis dates back to 1992, where Jack et al.[JPOT92] have proposed a model to estimate hippocampus volume atrophy from sMRI to assist in the clinical diagnosis of AD subjects . Since then,

automatic AD diagnosis based on hippocampus volume estimation has evolved with advances in neuroimaging and the advent of more advanced deep learning techniques.

The majority of the research studies that utilized deep learning methods for AD classification falls into three main categories: The first group utilize CNN models for feature extraction followed by a classifier such as support vector machines (SVM) [SS13, SLS+14b]. The second group relies on generated hand-crafted features (e.g., voxel-based or region-based) from neuroimaging modalities followed by the use of CNN-based classifiers. These models particularly apply other tools or algorithms to perform brain template mapping and multi-atlas (affine or rigid) registration approach to perform the feature extraction process. FreeSurfer and FIRST are two software suites which are heavily used for this purpose. In some cases, the ROI is defined by experts. These feature sets are utilized to train a CNN to classify the subjects [LLC+14, TAE+20, TAS+19, ATAA18, TAS+18]. These models are highly prone to registration errors and the presence of noise. Furthermore, the trained classifiers could be biased by the specific dataset used in trained phase. These methods usually result in a sub-optimal performance on disease diagnosis. The last group consists of models that have an end to end pipeline capable of direct feature extraction integrated with a classifier to model the feature space. These models are relatively newer and more powerful than their earlier counterparts [HAKEB16, ZGG+16, LLZS18, KSBD17]. To create such a pipeline, this study focuses on modeling the automatic hippocampus segmentation as a two-stage model relying on a localization process that delineates an area expected to contain the hippocampus structure, followed by a segmentation process. Subsequently, volumetric measures of disease prone areas such as the hippocampus could be used as reliable measures that would augment the prospects for multimodal and multiclass

classification and prediction algorithms.

To diagnose AD subjects based on the hippocampus structure, Li et al. [LLI+19] have proposed a hybrid convolutional model cascaded by a bidirectionally-gated Recurrent Neural Network(RNN). They have used a DenseNet-based model on 3D segments of the brain hippocampus to learn about related shape features and intensity. This CNN is followed by an RNN to learn high-level features of the Hippocampus for the diagnosis of AD subjects. While the results show the superiority of their proposed model on segregation of the AD subjects, their model highly depends on manual pre-processing and using external software for brain segmentation through brain registration. This pipeline is not only prone to noise of the external software, but also needs a high degree of manual scrutiny and data integrity checks as pre-processing steps.

A pre-trained 3D convolutional auto-encoder presented in [HAKEB16] was used to detect anatomical shape changes in sMRI. Few deeper fully connected layers of this model are fine-tuned for AD task-specific classification. Suk et al. also utilized an auto-encoder to extract features from PET, MRI and cerebrospinal fluid (CSF) separately and combined those features with AD Assessment Scale-cognitive (ADAScog) scores and clinical Mini-Mental State Examination (MMSE) to train a multi-kernel SVM to classify the subjects [SLS+14a].

5.3 Methods and Network Architecture Design

In this section, we introduce a novel framework for hippocampus segmentation. This framework is composed of two modules: 1) hippocampus localization, and 2) hippocampus segmentation. In the first module, a model estimates the hippocampus location in the brain and produces a cropped area surrounding the hippocampus

tissue. In the second module, the cropped area expected to contain the hippocampus structure is fed to a segmentation model. The design of the first module is inspired by [LZAS18, BCL+19, HA17]. This is a heuristic algorithm that performs 3D skull stripping and extracts the brain volume. Considering the ratio of the acquired volume to the average of the training set volume it performs a relative distance estimation, which defines the distance of the first slice of the brain to the first slice where the hippocampus appears in all three views of coronal, sagittal and axial to determine a rough estimate of the hippocampus location in the 3D MRI.

In the second module, the cropped area which is expected to contain the hippocampus is fed into a transformer-based segmentation model. The segmentation algorithm leverages the architecture design of TransUnet proposed in [CLY+21]. This new design is based on integrating the Vision Transformer and the UNet model, which has shown promising segmentation results on abdominal CT scans, but has not been explored for 3D brain MRI segmentation. This latter task is viewed as even more challenging given the difficulty in delineating the different regions of the brain in 3D MRI.

The hippocampus segmentation problem falls in the category of an imbalanced segmentation as the proportion of the region of interest is less than that of the background. To improve the original implementation of the model for handling this imbalanced case, we have used a combined loss function based on the focal loss idea. Some data augmentation techniques applied on the data (random rotation and flipping) before feeding the data points into the segmentation model. The proposed pipeline is depicted in figure 5.1.

Several interesting methods are proposed for strengthening the UNet architecture by coupling it with different backbone methods to extract richer features for more accurate segmentation. DenseNet, InceptionNet and ResNet when integrated

to the UNet architecture have resulted in higher performance in the different segmentation tasks [LCQ+18, DAD18, CRBFS19, QJZ+20, DÇ20, ZCN+17]. Following this line of research, we also explored multiple powerful backbone architecture that have been lately introduced for generating rich feature maps but have not been yet exploited in medical application. As such, networks like SeNet and EfficientNet are integrated into UNet and UNet++ to further improve the segmentation accuracy. This investigation has been performed to analyse the effect of more advanced feature extraction backbones on the well performing segmentation models such as UNet++. These results confirm that a high performing feature extractor does lead to a better performance of UNet type architecture overall, however they still lag behind the performance attained when integrating the Vision Transformer to the UNet architecture.

Furthermore, we have explored a multi-scale attention based network called MANet. This attention-based architecture has been proposed in [FWLW20] for liver tumor segmentation task, but has not been applied for brain segmentation so far.

An overview of the studied networks and relevant details of model design and evaluation metrics are discussed next.

UNet architecture which was proposed in 2015 by Ronneberger et al. [RFB15], has been one of the most dominant method in medical image segmentation. Since then this model has served as a building block of many other image segmentation models. In contrast to object detection which draws a bounding box around the subject and defines its corresponding label, in image segmentation, a fine binary map draws over the images and classifies each pixel separating background and object/region of interest.

UNet architecture is composed of two main paths. Multiple convolution layers are followed by a max-pooling layer from the encoding path. Through this path, the model learns spatially relevant contextual information. This path matches a reverse decoding path which adds precise localization to yield a final segmentation with same size as the input image.

UNet++ that was proposed in 2018 attempts to improve the UNet architecture by re-designing the encoder and decoder path and skip connections of the original UNet. These pathways in UNet++ are composed of a series of nested dense connections that reduce the semantic gap between encoder and decoder's feature maps. This strengthened connections in Unet++ show considerable improvement in segmentation tasks [ZSTL18].

DenseNet is an effective network which consists of several dense blocks. Inside each dense block, every layer's output directly connects to all the following layers. Therefore the input to each layer is concatenated to all the subsequent feature maps. This design gives the model several advantages: 1) It prevents the vanishing gradient by constructing a larger, richer feature map, 2) having multiple 1*1 kernels in the network as means to control the computational complexity of the model,and 3) having all the features from the input fed to the final layer instead of only focusing on final layer features, all of which lead to higher model performance [HLVDMW17].

EfficientNet was introduced by Tan et al. in 2019 to show how systematically scaling and balancing network depth, width and resolution can improve the performance. This new design leads to a family of models called EfficientNet-B7 with a better accuracy than previous ConvNets while they have much smaller size (smaller by a scale of 8.4) and lower inference time(6.1x faster) [TL19].

In contrast to many CNN model-based architectures which attempt to enhance

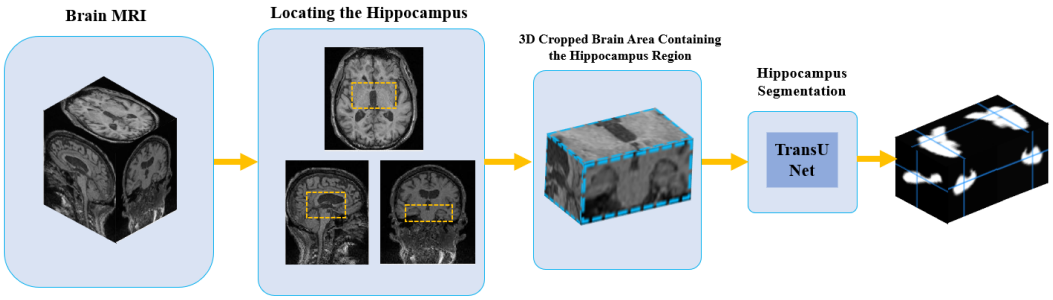

Figure 5.1: The two-stage proposed pipeline for hippocampus segmentation in MR images

spatial encoding quality throughout the feature hierarchy to achieve better representational power, Squeeze and Excitation network (SeNet) focuses on channel relationship [HSS18]. A squeeze and excitation block is introduced to explicitly model inter dependencies between various channels. This action results in calibrating channel-wise feature response to improve the network power. This new architecture not only generalizes well to other applications but also improves the CNN performance significantly .

Attention mechanism was first proposed for natural language processing task and more recently expanded to the image processing and computer vision domains. Attention mechanism draws from human vision, in that once we know the context in which an object appears in a scene, we look for that same context when we search for that object in the future. Multiple research endeavors have improved their design by adding attention mechanism in conjunction with convolution layer, one of them is MANet. While many of the proposed UNet architectures are based on multi-scale feature fusion, MANet suggests a new attention-based model. This network first introduced for liver tumor segmentation task and has a self-attention mechanism which integrates local features with associated global dependencies. It is composed of two main blocks:(1) a position-wise attention block which tries to

find and model the spatial dependencies between pixels in a global view, and (2) a multi-scale fusion attention block which applies multi-scale semantic feature fusion to model the channel dependencies between feature maps [FWLW20].

While there are several studies which have exploited attention mechanism for image classification, a fully transformer-based model has been proposed by Google research team more recently in late 2020. This architecture is identical to the original transformer model proposed for Natural Language Processing (NLP). It processes a sequence of image patches similar to NLP tokens for image classification task. Vision Transformer has shown promising results compared to the state-of-the-art CNN models if it is trained on a large dataset for enough time while requiring substantially fewer computational resources to train [DBK+20].

This novel transformer-based architecture is applied for image classification purposes using an encoding module. For each image that is processed the model predicts a label. To make this model applicable to more complex tasks such as object detection and image segmentation, some modification in the architecture is essential. To apply this model for segmentation task, Chen et al. have coupled this architecture with a decoding module inspired by UNet architecture.They also applied the transformer encoding on the feature maps extracted from layer third of a ResNet50 network. They selected this design after failing to obtain compelling result by following solely the original architecture which directly tokenize the original image. This CNN-Transformer hybrid design performs better than pure transformer encoding as it allows the network to exploit high resolution CNN feature maps in the reconstruction path. The reconstruction path consists of several up-sampling units. The very first reconstruction unit gets the output of the transformer encoder and after up-sampling it, it concatenates the current feature map with the feature map of the last CNN layer of the ResNet in the corresponding encoding path to

incorporate multi-scale information into the model. The outcomes passes through a 3*3 convolution and ReLU layer to form the input for the next reconstruction unit. Same process applies two more times on the resulting output of each layer until the decoder reconstructs the segmentation task in the original size of the input. A segmentation head is added at the end of the reconstruction path which classifies each pixel to its corresponding class and recovers the segmentation mask with the same resolution of the input image [CLY+21].

The network architecture which is adopted from Trans+Unet model is depicted in figure 5.2

To have a better performance in an imbalanced segmentation task, we have changed the original loss function to a combined focal loss inspired by [YSSR21].

In our design, unlike in the original Vision Transformer model, the image is passed through a CNN model to generate a rich feature map. Furthermore the first few intermediate feature maps in the ResNet module are also kept to help reconstruction in the up-sampling path. The final feature map which has a 2D shape will split into fixed-size 1*1 patches. Patches are flattened and linearly projected to a new latent space. To retain positional information, position embedding are added to each patch separately as input to the transformer encoder unit. the Transformer layer consists of layer norm and Multi Head Attention (MHA) unit. In this model, 12 transformer units have been stacked on top of each other. The final feature map is bi-linearly up-sampled and concatenated with the corresponding feature map in the down-sampling path from CNN model. Each up-sampling block consists of a 2 up-sampling operator, a 3×3 convolution layer, and a ReLU function.

Considering the enormous success of the aforementioned feature extractor networks (ResNet, DPN ,SeNet, EfficientNet, ExceptionNet) in many segmentation tasks, we have integrated them into our pipeline and study them extensively when

they are paired with 3 distinct segmentation models (UNet,UNet++,MANet) to evaluate their performance when dealing with highly imbalanced segmentation tasks with the specific challenges such as when dealing with convoluted structures like the hippocampus and low contrast margins in between the different brain regions.

5.3.1 Loss Function

Class imbalance is a major issue that impact the semantic segmentation task, especially in neuroimaging. Since small ROI is usually suppressed through max pooling layers, solutions based on optimizing the cross entropy loss function are often unsatisfactory. To overcome this issue, we have adopted mixed Focal Dice loss function. This loss function is a weighted combination of modified focal loss and modified focal dice loss. Focal loss was first introduced to address the problem of class imbalance faced by cross entropy loss. To do so, it down-weights the contribution of easy examples which in turn enables learning from harder examples. In this study, as we face class imbalance in this segmentation problem, we have adopted a weighted combination of modified focal loss L_{mF} and modified focal dice loss L_{mFD} as shown below:

$$L_{MF} = \lambda L_{mF} + (1 - \lambda) L_{mFD}$$

$$L_{mFD} = \sum_{c=1}^{C} (1 - mD)^{\frac{1}{\gamma}} (2)$$

$$L_{mF} = -\alpha (1 - p_t)^{\gamma} . L_{mCE} (3)$$

$$L_{mCE} = -\frac{1}{N} \sum_{i=1}^{N} \beta (t_i - log(p_i)) + (1 - \beta) \left[(1 - t_i) ln(1 - p_i) \right]$$

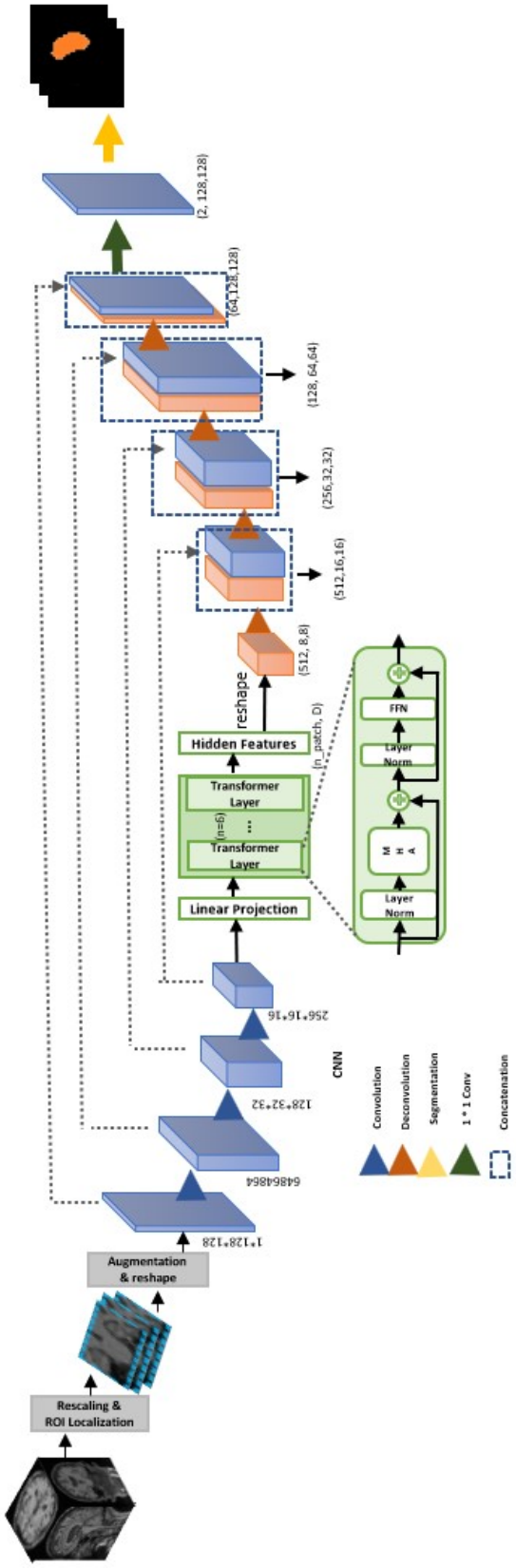

Figure 5.2: Detailed illustration of the proposed architecture

$$p_t = \begin{cases} p & \text{if y} = 1 \\ 1-p & \text{if y} = 0 \end{cases}$$

where $\lambda \in [0, 1]$ defines the relative weights of two components of the loss function and γ is the focal parameter. Parameter C defines the number of classes,

The t_i in the L_{mCE} equation refers to the Tversky index as an asymmetric similarity measure, which is closely related to the Dice score and enables optimization for output imbalance by tuning the weights assigned to false positives and false negatives. The t_i calculation is defined with details in [YSSR21]. The α term in the range of [0, 1] controls the relative weighting of the Dice and cross entropy terms contribution to the loss, and β controls the relative weights assigned to false positives and negatives. A value of $\beta > \frac{1}{2}$ penalizes false negative predictions more so than false positives.

5.4 Evaluation Metrics

To quantitatively evaluate and compare the performance of the proposed methods, four metrics have been exploited. Mean Dice Similarity Coefficient (DCS) is used to measure overlaps between ground truth mask A_g and predicted mask A_p.

$$DSC = \frac{1}{n} \times \sum_{i=1}^{n} \frac{|A_{s_i} \cap A_{g_i}|}{|A_{s_i}| + |A_{g_i}|}$$

Jaccard Similarity Coefficient (JSC) is adopted to compare the similarity and diversity between A_g and A_p.

$$JSC = \frac{1}{n} \times \sum_{i=1}^{n} \frac{|A_{s_i} \cap A_{g_i}|}{|A_{s_i}| + |A_{g_i}| - |A_{s_i} \cup A_{g_i}|}$$

Precision Index shows the overlapping ratio between A_g and A_p over ground truth mask A_g while Recall Index (RI) shows the overlapping ratio between A_g and A_p over predicted mask A_p.

$$Precision\ Index = \frac{1}{n} \times \sum_{i=1}^{n} \frac{|A_{s_i} \cap A_{g_i}|}{|A_{g_i}|}$$

$$Recall\ Index = \frac{1}{n} \times \sum_{i=1}^{n} \frac{|A_{s_i} \cap A_{g_i}|}{|A_{s_i}|}$$

All these metrics are calculated per image and the mean of all the metrics over the test dataset is reported. A good segmentation method should produce high value in all the metrics.

5.4.1 Optimizer Algorithm

One of the most popular optimization algorithms in machine learning domain is Gradient Descent. It is an iterative algorithm looking for global minimum of a differentiable function. In each step, GD calculates the derivative of the loss function which points to the direction of the steepest descent. Various modification of this algorithm have been proposed to improve the calculation speed and prevent the algorithm to stuck in local minima.

Till now, several other optimization algorithms have been proposed to improve GD peformance in terms of speed and convergence rate, including Stochastic Gradient Descent (SGD) [Bot10] , AdaGrad [DHS11], RMSProp [TH$^+$12], and Adam [KB14], to name a few. Adam combines AdaGrad and RMSProp to create a better optimization algorithm. Adam stands for Adaptive Moment Estimation. Similar to momentum, this algorithm has an exponentially decaying average of past gradients. So the direction of weight updates is calculated in a similar manner to momentum.

Adam also employs an exponentially decaying average of past squared gradients to support an adaptive learning rate [KB14].

5.4.2 Activation Function

Activation function is a non linear function in a neural network neuron that delivers an output based on inputs. Sigmoid activation function is one of the earliest non linearity which has been used in CNNs. Several other non-linearity functions have been proposed till now, including ReLu, Leaky Relu, Selu, Gelu and Mish, each has own it's advantage and weaknesses.

Among all,Mish has outperformed other vastly used activation functions like Relu and Swish in more than 70 challenging problems. Mish is a non-monotonic and smooth activation function which has several properties that can improve the model performance when compared to the well studied and popularly used activation functions like Relu, LeakyRelu or Swish. Main properties of Mish are concluded below: $f(X) = X.tanh(ln(1 + e^x))$ Mish

- Unbounded above and bounded below: Unbounded above is a desirable property for any activation function since it avoids saturation which causes the training to slow down drastically. Hence, speeding up the training process. The bounded below property helps in achieving strong regularization effects (fits the model properly). (This property of Mish is similar to the properties of ReLU and Swish with a range [0.31,)).
- Non-monotonic function: This property helps preserve the small negative values, hence stabilizing the network gradient flow. Most commonly used activation function like ReLU [f(x) = max(0, x)], and Leaky ReLU [f(x) = max(0,

x), 1] fail to preserve the negative values as their differentiation is 0, and hence most of the neurons do not get updated.

- Infinite order of Continuity and Smooth: Mish being a smooth function is good with the improvement of results as it is better at generalization and effective optimization of results.
- High computational cost, but better performance: It is costly as compared to ReLU but shows better results in the deep neural networks as compared to ReLU.
- Self Gating: This property is inspired by Swish function, where the scalar input is provided to the gate. It is advantageous over point-wise activation functions like ReLU which take in a single scalar input without requiring to change the network parameters.

Weight initialization have done with Zhavier technique. Batch normalization layer employed after all the non linearity layers. The method has been tested on 251 T1 MRI images each having 250 to 350 slices.

5.5 Dataset

In this study we have used two datasets. The data for training the segmentation model has been captured from the Medical Segmentation Decathlon Challenge, 2019. It consists of 105 patient data and 90 healthy subjects. The 3D structural MRI data were captured by a 3D T1-weighted Magnetization Prepared Rapid Gradient Echo Imaging (MPRAGE) sequence (TI/TR/TE, 860/8.0/3.7 ms; 170 sagittal slices; voxel size, 1.0 mm3) all with the same machine. Tracing of head, body, and tail of the hippocampus have been performed on the entire data [SAB$^+$19]. The data

consisting of 326 MRIs used to test the pipeline end to end has been acquired from Mount Sinai Medical Center (MSMC), Miami Florida as part of the data for the 1Florida Alzheimer's Disease Research Center (ADRC). We first apply the MR-based skull-stripping technique to extract the brain from each MRI scan. Then the hippocampus is segmented in each brain image separately. The volumetric results from this second dataset is compared with the volumetric results obtained using FreeSurfer 6.0 (see supplementary information).

5.6 Experiments and Results

The experiments were carried out on an NVidia Titan RTX with 576 Tensor Cores for AI acceleration and 24 GB of GDDR6 (Graphics Double Data Rate 6) memory. The data has been normalized prior to running the experiments. With the Decathlon Challenge dataset, the segmentation dataset was divided into two distinct subsets: 194 (3D) volumes for training and validation through cross validation technique and the rest, the 65 (3D) volumes for testing. The batch size is set to 24 and the number of training epoch is set to 150. During every training iteration, the input was augmented by random rotation and flipping. The training process took about 13 hours. Five folds cross validation technique has been deployed to train and test the end to end pipeline. For the MSMC dataset, we used 326 MRI sets and deployed our pipeline on them.

Optimization of the deep learning model is achieved using the Adam optimizer with a dynamic learning rate of 10e-4. Although the first component in the proposed architecture is designed to reduce the imbalance between ROI and background by cropping a region in the brain expected to contain the hippocampus. However, due to the small size of the hippocampus in comparison to the background in the cropped

3D (1:5), the data is still considered imbalanced. To tackle this issue, a combination of loss functions [SLV+17] is used for this purpose. Focal dice loss and focal loss have been utilized as they help when dealing with imbalanced data, and a cross entropy (CE) function is used as means to gauge the difference between the predicted value from the true label. The choice of the CE function is due to its smooth nature and the easiness for calculating the gradient. Similar to [YXW+18, ZXS+17], 3D volumes are processed slice by slice, and in the test time, the final 3D structure is reconstructed by stacking all the slices together.

Evaluation of the proposed method was done by the mean Dice similarity coefficient, Jaccard score, mean precision, and mean recall. Related equations for calculating these metrics were provided earlier in the Evaluation Metrics section. To ensure a comprehensive and fair comparison, we have coupled different models with multiple backbone architectures. The change in receptive field in the various design architectures along with the different choices of loss functions, depth and width of network and other design details resulted in the need for different segmentation masks.

The performance of the proposed two-stage architecture for carrying out the hippocampus segmentation task has been compared with various state of the art segmentation network including Vanilla UNet, and UNet++. The implementation of these networks has been borrowed from the PyTorch segmentation models library [Yak20].

For illustrative purposes, figure 5.3 shows the model segmentation mask against FreeSurfer mask and manual segmentation mask in three different subjects with the axial, coronal and sagittal views. These results for each slice are then combined to generate the 3D volumes of the hippocampus region for each subject.

Table 5.1 provides a quantitative comparison of the segmentation performance

of UNet network coupled with several outstanding feature extractor backbone architectures. Except for DenseNet and ResNet, this is the first study that couples most advanced feature extractors like SeNet, EfficientNet and ResNext with UNet. As can be seen, UNet-SeNet and UNet-ResNext combinations have the highest DSC and JSC values in the current segmentation task, which supports the fact that SeNet and ResNext are stronger feature extractors than ResNet or the original UNet backbone architecture.

Table 5.2 summarizes the results of similar studies when coupling state-of-the-art feature extractors with the UNet++ segmentation model. These measured metrics are slightly higher values than their counterpart in table 5.2 due to the better performance of UNet++ with almost all the combined feature extractors. UNet++ is an upgraded version of UNet with dense skip pathways that attempt to reduce the semantic gap between the feature maps of the encoder and decoder modules. From this table, the same conclusion that SeNet and ResNext do enhance the segmentation model can be drawn.

Table 5.3 shows the result of an experiment with MANet segmentation model paired with a similar set of backbone architectures. Considering that MANet has outperformed UNet++ in kidney segmentation task, it shows relatively similar outcomes to UNet++; however, not all backbone architectures increase the segmentation results as in UNet++. This difference can be explained by the attention-based mechanism of MANet, which is not a typical CNN based architecture.

In our research, we have searched through the best segmentation model for the second stage of our pipeline. We have investigated various segmentation models such as UNet and UNet++ which are based on convolution mechanism, and the more recent architectures which are based on a combination of convolution and attention mechanism such as MANet. The results presented in table 5.4 and figure 5.4 shows

Table 5.1: Performance of UNet model paired with different feature extraction backbones. When a reference is not available, * means implemented by our group

Method	Backbone	Dice	Jaccard	Precision	Recall	Ref
UNet	ResNet	0.882	0.791	0.883	0.884	[ZCN+17]
UNet	DenseNet	0.885	0.795	0.893	0.878	[LCQ+18]
UNet	ResNext	0.885	0.796	0.912	0.862	*
UNet	EfficientNet	0.886	0.797	0.879	0.895	*
UNet	Xception	0.884	0.794	0.893	0.877	*
UNet	**SeNet**	**0.888**	**0.801**	**0.896**	**0.883**	*****

Table 5.2: Performance of UNet++ model paired with different feature extraction backbones. When a reference is not available, * means implemented by our group

Method	Backbone	Dice	Jaccard	Precision	Recall	Ref
UNet++	ResNet	0.889	0.802	0.899	0.882	[JSR+19]
UNet++	ResNext	0.893	0.802	0.8992	0.882	*
UNet++	EfficientNet	0.8865	0.797	0.879	0.895	*
UNet++	Xception	0.886	0.801	0.897	0.879	*
UNet++	**SeNet**	**0.893**	**0.809**	**0.899**	**0.889**	*

that our proposed model performs at least 2% higher in all four metrics of DSC, JSC, PI and RI in comparison to UNet, UNet++ and MANet.

The best performing segmentation-feature extraction pairs from the above experiments have been gathered and compared together in figure 5.4. As can be seen from the results, the best performance is attributed to Transformer-Unet with combined loss functions. This network is one of the most recent attention-based model with Vision Transformers and Resnet feature extractor combined. The result in this figure shows better performance of the Transformer-Unet model over other models as expected in both metrics of DSC and JSC. This can be explained by the high performance of the Vision Transformer in terms of feature extraction. Finally we have performed Anova test to verify the statistical significance of the difference in the model performance (P-value greater than 0.05)

Table 5.3: Performance of MANet model paired with different feature extraction backbones. All combination have been proposed by our group

Method	Backbone	Dice	Jaccard	Precision	Recall
MANet	ResNet	0.886	0.796	0.896	0.877
MANet	ResNext	0.801	0.717	0.819	0.792
MANet	DenseNet	0.877	0.785	0.920	0.841
MANet	EfficientNet	0.881	0.790	0.903	0.862
MANet	Xception	0.872	0.777	0.921	0.832
MANet	**SeNet**	**0.891**	**0.805**	**0.897**	0.887

Table 5.4: The Performance of different segmentation methods over Dechathlon dataset. *Results of best performing combination of backbone for UNet,UNet++ and MANet models are illustrated.

Networks	Dice	Jaccard	Precision	Recall
UNet	0.888	0.801	0.896	0.883
UNet++	0.893	0.808	0.899	0.898
MANet	0.891	0.805	0.897	0.887
Transformer based	**0.911**	**0.821**	**0.923**	**0.913**

5.7 Discussion

Hippocampus is one of the most important disease-prone region of the brain used as a biomarker for detecting the onset of AD as well as for assessing other neurological, neuro-psychological and neurodegenerative diseases like epilepsy, depression, bipolar disorder, schizophrenia, and Parkinson, to name a few . To assist doctors in analyzing patients neuroimaging data, automatic, easy to use and deploy, fast and accurate segmentation models play an important role for making a correct diagnosis. Therefore in this study, we proposed a new two-stage pipeline for segmentation of the hippocampus tissue in the 3D brain MRI. This pipeline is adopted mainly to handle the imbalance in data which results from small size ROI like the hippocampus

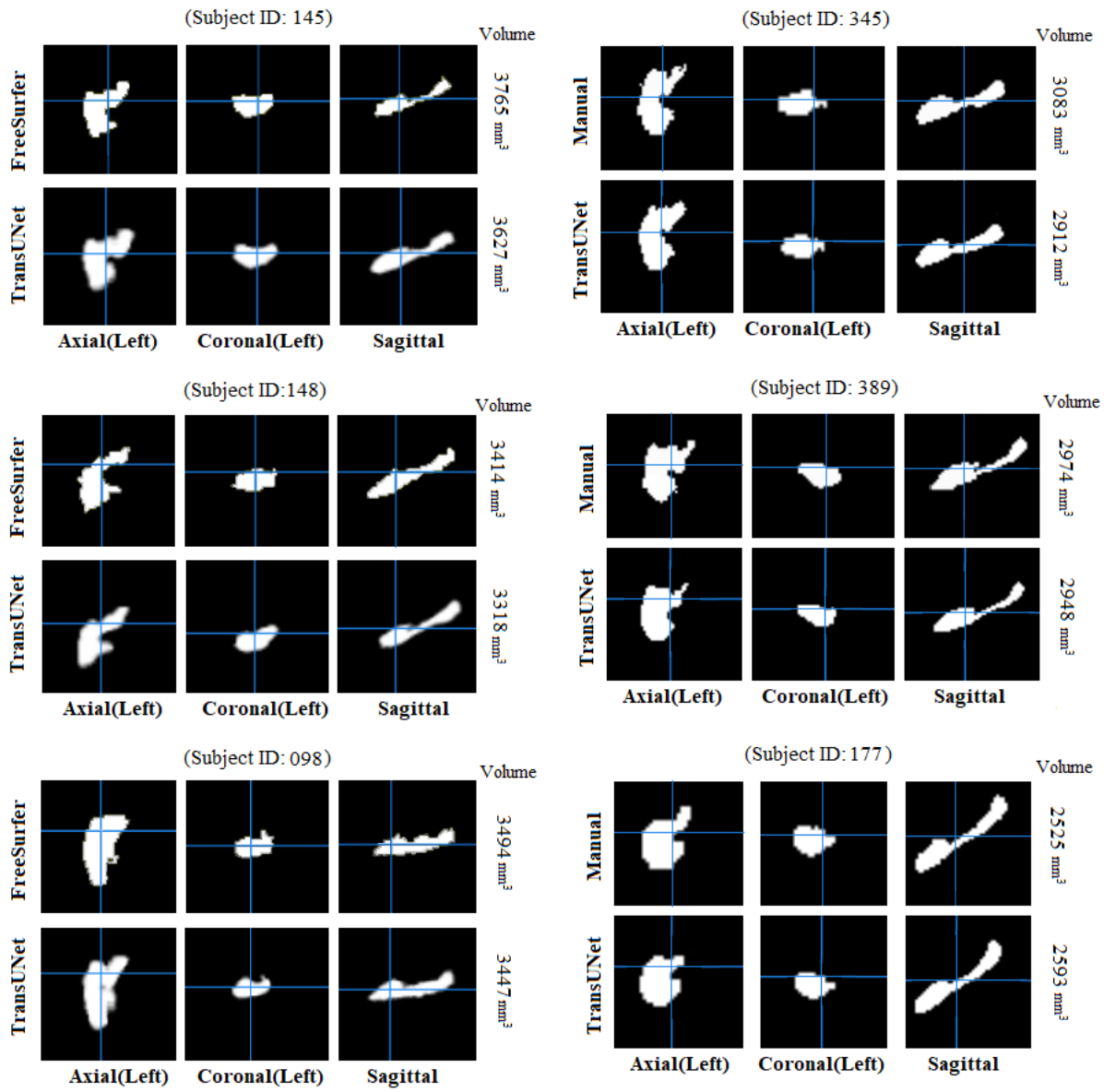

Figure 5.3: Illustration of model segmentation against FreeSurfer and manual segmentation masks in different subjects has shown in left and right columns respectively. Each subject hippocampus has been visualized in axial, sagittal and coronal view

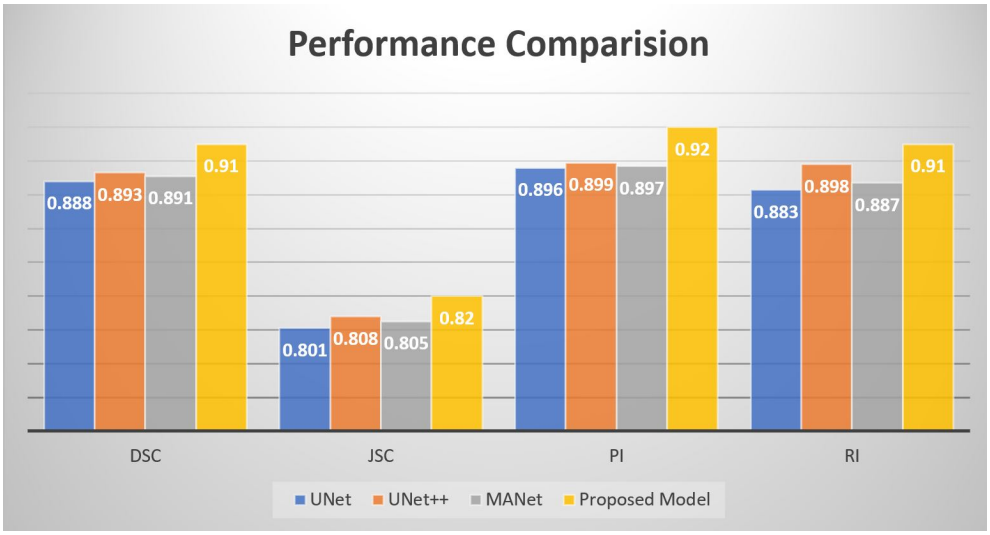

Figure 5.4: Performance comparison of top performing models integrated in proposed two-stage pipeline on hippocampus segmentation task.

in contrast to the larger background. The first module follows an intuitive approach that roughly estimates the location of the ROI in the brain MRI and makes a crop around that region. More specifically, first module starts with skull striping and equalizing slice spacing in all 3 direction. Then it defines the coordinates of the last slices that hippocampus occupies with respect to the brain atlas, then it crops the image and sends the cropped section to the segmentation module. We have strengthened the segmentation module by incorporating the advantages that vision transformers provide. The vision transformer has been recently released by Google and is trained on very large datasets and has shown tremendous success in feature extraction, surpassing the already powerful ResNet performance. We further investigated the power of UNet, UNet++ and MANet for the current task when coupled with stronger feature extraction backbone architectures such as EfficientNet, SeNet or Xception. Our results support the assumption that a stronger backbone architecture leads to a better segmentation performance, while Transformer-Unet with an adjusted loss can improve these results further. Our proposed pipeline based on

Transformer-Unet with the new combination of dice and focal loss functions is found to be more suited for the aforementioned imbalance problem in brain segmentation. This new pipeline has shown over 2% and 1% improvement in Dice and Jaccard Coefficient Similarity Score, respectively.

Through this research endeavor, the following contributions are made: 1) Investigated and extended the segmentation models provided by previous studies which a focus on enhancing UNet architecture by coupling it with a stronger backbone architecture for hippocampus segmentation. For the first time, a more advanced UNet architecture is proposed with better performance than original UNet and other networks such as UNet-SeNet and UNet++-SeNet. 2) Proposed an integrated two-stage pipeline to locate and then segment the hippocampus structure, while addressing the unequal distribution of foreground and background elements. 3) Leveraged a Vision Transformer-based architecture coupled with the UNet architecture trained for hippocampus segmentation task by adopting a modified focal dice loss function which is shown to optimize training on imbalanced data. 4) Compared the segmentation results obtained with the manual segmentation provided through the Medical Segmentation Decathlon Challenge of 2019 on 40 MRIs not seen in the training phase, showing a mean volume difference of 5% between them with a standard deviation of 3%. 5) Deployed the proposed segmentation method over our own Mount Sinai data with the 1Florida Alzheimer's Disease Research Center (ADRC) data, consisting of 326 MRIs. A comparison with the FreeSurfer version 6.0 results showed a mean volume difference of 7 % with a standard deviation of 4%. 6) Performed an exhaustive comparative assessment of relevant image segmentation architectures like UNet, UNet++ and MANet coupled with stronger feature extraction backbone architectures such as ResNet, DenseNet, EfficientNet, SeNet or Xception using the metrics of Dice, mean Jaccard, Precision, Recall and average

symmetric surface distance (ASD).

In terms of limitations, when deploying the proposed algorithm to process a different dataset as we did with the Mount Sinai data, the learned features of the base network will have to be repurposed to the target domain to ensure a good performance. To transfer knowledge from the source domain, the pre-trained network structure can thus be fully or partially utilized for the new task at hand. Also, using the new labeled data, the network can be adapted and re-trained on the new dataset that has been transformed to the same shape, slicing size, and dimensions as the original dataset on which the original model was developed.

CHAPTER 6

CONCLUDING REMARKS AND FUTURE WORK

6.1 Concluding Remarks

In this dissertation, the impact of multiple imputation techniques have been investigated for estimating the missing values. Various imputation algorithms were coupled with three best performing classification algorithms to improve the accuracy of the multiclass classification of AD subjects between four groups of AD, EMCI, LMCI and Control Normal. In another attempt to exploit subject's longitudinal data, two modifications of the Recurrent Neural Networks: Long Short Term Memory and Gated Recurrent Network, which are known to be very effective in finding the pattern in sequential data, were deployed to ascertain their efficacy in the presence of missing data. Several approaches for handling the missing data in those models have been investigated.

When addressing the challenge of incomplete multimodal datasets for Alzheimer diagnosis, an extensive comparison of methods for estimating missing values in large heterogeneous dataset was provided. The inherent challenge of missing values in longitudinal and multimodal studies is compounded in complexity by the heterogeneity of the data. Under the missing data challenge, the Gradient Boosting algorithm is found to yield the highest performance when dealing with multiclass classification as a more natural process when dealing with all subgroups and the different prodromal stages of the disease. An advantage of the research presented in Chapter 3 is in the broad scope of the investigation which coupled different classifiers with the four most relevant imputation techniques including the k-nearest neighbors impute algorithm (KNN), Matrix Factorization, SVD, and Soft Impute. Findings suggest that imputation techniques perform reasonably well when there is a low percentage

of missing values in a given dataset, and that all these imputation techniques fail to perform well when high levels of missing data are experienced. Moreover, when a particular modality is completely missing for so many subjects within the dataset, a situation we refer to as block-wise missing data patterns, the imputation methods fail to make any statistically meaningful inferences on the existing data to estimate the missing ones.

Considering the importance of the hippocampus as a disease prone region for many neurological and neurodegenerative diseases, an another research endeavor of this dissertation involve the development of a deep learning model for the localisation and segmentation of the hippocampus region as a biomarker for AD. The developed model was composed of two stages. In the first stage a hippocampus localization model tries to detect a 3D bounding box around the hippocampus tissue and in the second stage a segmentation model defines the accurate boundaries of the hippocampus tissue. This two-stage model further strengthened by a combined loss function to address the imbalance in the data, as the hippocampus tissue is extremely small in context to the 3D full brain MRI. The small ratio of the ROI to background creates an imbalanced segmentation problem, where the majority of the image pixels belong to the negative class. The second stage of the model which is the segmentation module combines two well-performing deep learning frameworks: UNET and Vision Transformer. Effects of well known activation functions along with a few successful new versions such as MISH activation function have been tested and reported. Final model performance against original UNet model and several variants of UNet (UNET++, UNetSenet, UNetDenseNet,etc) were summarized in this chapter. An extremely important development in this case is the fact that the developed algorithm can be deployed to other datasets as we have done with the Mount Sinai Medical Center data as part of the 1Florida ADRC, show-

ing extremely important results that are compared to the well-known FreeSurfer software suite.

6.2 Future Research

Trajectory projection of Alzheimer's Disease (AD) progression has been halted for a long time due to lack of sufficient longitudinal data. In recent decades ADNI has realized a relatively large dataset of AD subjects longitudinal studies which enables researchers to focus on progression modeling of the disease. Alzheimer's Disease progression is generally assessed using biomarkers including structural Magnetic Resonance Imaging (MRI), 18-Fluoro-DeoxyGlucose Positron Emission Tomography (FDG-PET) imaging,CerebroSpinal Fluid (CSF), cognitive examination, and to alesser extent electroencephalography (EEG)[JPM18, PDHvdF$^+$13, LCD$^+$18]. However several studies in the literature only focus on the effect of single biomarker or only one modality for diagnosis and/or prognosis of the disease. When contending with the missing data challenge inherent to longitudinal and multimodal studies, researchers need to situate this challenge in context to the more difficult process of multiclass classification. The low accuracy obtained when performing multiclass classification and the failure of imputation methods due to large percentage of missing values or block-wise missing data is due to the variability in measurements used and to the interrelatedness between them. The cognitive tests used to label subjects at baseline must be augmented through specific neuroimage measurements that are yet to be defined through the most relevant feature extraction methods.

In relation to the results obtained in Chapter 3, future work should focus on improving the current multiclass classification accuracy with the application of newer techniques such as the Optimal Margin Distribution [ZZ19]. Also developing newer

machine learning classification methods that could find a balance in the features that are optimal not only for multiclass classification but also for prediction as well, and to determine a breakpoint when optimizing one task could only be achieved at the expense of another task, especially when we have to contend with in incomplete datasets.

In this dissertation, the applications of LSTM and GRUs to model prediction tasks over the longitudinal data from the ADNI dataset were introduced. The proposed models can be used for the diagnosis of Alzheimer's disease. We also incorporated three different strategies to deal with the incomplete and missing data (from time points and modalities). Trying different variations of RNNs (i.e., LSTM and GRU), We found slightly better performance using the LSTM model. Our model can classify AD vs.NC with an accuracy of 95.9%, even with simple replicate and zero filling of the missing data. It also performs better classification of AD vs.MCI and NC vs.MCI patients. As a direction for future works, designing an end-to-end convolutional and LSTM model for this longitudinal dataset can be of great interest, to accurately learn powerful image features (from MRI and PET) and simultaneously learn the classifier parameters.

To better appreciate the results summarized in terms of mean volume difference and standard deviation as reported in 3) and 4) earlier in the Discussion section of , and in order to understand the complexity in segmenting brain regions like the hippocampus in 3D MRI, the study reported in [BBG+15] provides all the evidence we need in the challenge faced when segmenting brain regions in MR images. The authors of this study show that even when there is very high agreement among four expert tracers (pairwise Jaccard indices 0.82-0.87), the volumetric results among the four expert tracers obtained using the HarP benchmark dataset consisting of 135 MRIs still showed a mean volume difference of 9% between them with a standard

deviation of 7%. Note that in the results reported in Tables I through III, they show a Jaccard value between 0.80 to 0.82 for the different architecture with the best result obtained with the Transformer-based architecture. These results as obtained through the proposed machine learning method come in support of the call for automated means to segment disease prone regions like the hippocampus [JJAK$^+$11, WBP$^+$17, BBG$^+$15, BBM$^+$15].

For the future research, one can adopt this segmentation model to a pipeline to jointly extract features from MRI and other discriminative regions of the brain and classify the subjects in the subsequent network. More precisely, this end-to-end network architecture composed of a CNN-ViT model which extract features of the hippocampus from MRI. This network is jointly trained with another network which captures the features and models the subject space between four classes. The current heuristic hippocampus localization model can be improved if replaced by a trainable neural network to more accurately locate the bounding box around the hippocampus.

Moreover, it will be worth investigating ways to enhance the proposed modified TransUnet with a more powerful feature extraction process using image processing coupled with machine learning that exploits the relational positioning or location of brain anatomical landmarks like the ventricles and cerebellum. This process could also involve relational positioning in between brain regions. For example, once the hippocampus is extracted, the focus will shift on searching for the expected region containing the amygdala to be segmented as the next step. Image processing for enhancing these segmentation tasks could include histogram modifications that extend the dynamic range or perform histogram equalization followed by specialized edge detection methods [MRCA18] that would take into consideration any noise effect [MMC$^+$19]. Such imaging and machine learning algorithms could be further

integrated with the newest generation of Vision Transformers also called Token by Token transformers, which show considerable improvements over the traditional vision transformers for image classification and feature matching.

BIBLIOGRAPHY

[ADD+11] Marilyn S Albert, Steven T DeKosky, Dennis Dickson, Bruno Dubois, Howard H Feldman, Nick C Fox, Anthony Gamst, David M Holtzman, William J Jagust, Ronald C Petersen, et al. 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*, 7(3):270–279, 2011.

[AHH+09] Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Joseph V Hajnal, and Daniel Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.

[ATAA18] Maryamossadat Aghili, Solale Tabarestani, Malek Adjouadi, and Ehsan Adeli. Predictive modeling of longitudinal data for alzheimer's disease diagnosis using rnns. In *International Workshop on PRedictive Intelligence In MEdicine*, pages 112–119. Springer, 2018.

[BBG+15] Marina Boccardi, Martina Bocchetta, Rossana Ganzola, Nicolas Robitaille, Alberto Redolfi, Simon Duchesne, Clifford R Jack Jr, Giovanni B Frisoni, EADC-ADNI Working Group on The Harmonized Protocol for Manual Hippocampal Segmentation, for the Alzheimer's Disease Neuroimaging Initiative, George Bartzokis, et al. Operationalizing protocol differences for eadc-adni manual hippocampal segmentation. *Alzheimer's & Dementia*, 11(2):184–194, 2015.

[BBL+07] J Barnes, RG Boyes, EB Lewis, JM Schott, C Frost, RI Scahill, and NC Fox. Automatic calculation of hippocampal atrophy rates using a hippocampal template and the boundary shift integral. *Neurobiology of aging*, 28(11):1657–1663, 2007.

[BBM+15] Marina Boccardi, Martina Bocchetta, Félix C Morency, D Louis Collins, Masami Nishikawa, Rossana Ganzola, Michel J Grothe, Dominik Wolf, Alberto Redolfi, Michela Pievani, et al. Training labels for hippocampal segmentation based on the eadc-adni harmonized hippocampal protocol. *Alzheimer's & Dementia*, 11(2):175–183, 2015.

[BCL+19] Abol Basher, Kyu Yeong Choi, Jang Jae Lee, Bumshik Lee, Byeong C Kim, Kun Ho Lee, and Ho Yub Jung. Hippocampus

localization using a two-stage ensemble hough convolutional neural network. *IEEE Access*, 7:73436–73447, 2019.

[BDHN16] Ciprian D Billones, Olivia Jan Louville D Demetria, David Earl D Hostallero, and Prospero C Naval. Demnet: A convolutional neural network for the detection of alzheimer's disease and mild cognitive impairment. In *2016 IEEE Region 10 Conference (TENCON)*, pages 3724–3727. IEEE, 2016.

[BEWY] Seo-Jin Bang, CMU EDU, Yuchuan Wang, and Yang Yang. Phased-lstm based predictive model for longitudinal ehr data with missing values.

[BHR+16] Mark Belger, Josep Maria Haro, Catherine Reed, Michael Happich, Kristin Kahle-Wrobleski, Josep Maria Argimon, Giuseppe Bruno, Richard Dodel, Roy W Jones, Bruno Vellas, et al. How to deal with missing longitudinal data in cost of illness analysis in alzheimer's disease—suggestions from the geras observational study. *BMC medical research methodology*, 16(1):83, 2016.

[BKB+18] Andrew R Bender, Attila Keresztes, Nils C Bodammer, Yee Lee Shing, Markus Werkle-Bergner, Ana M Daugherty, Qijing Yu, Simone Kühn, Ulman Lindenberger, and Naftali Raz. Optimization and validation of automated hippocampal subfield segmentation across the lifespan. *Human brain mapping*, 39(2):916–931, 2018.

[BNA+00] J Douglas Bremner, Meena Narayan, Eric R Anderson, Lawrence H Staib, Helen L Miller, and Dennis S Charney. Hippocampal volume reduction in major depression. *American Journal of Psychiatry*, 157(1):115–118, 2000.

[Bot10] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[BRSM10] Mohd Ali Balafar, Abdul Rahman Ramli, M Iqbal Saripan, and Syamsiah Mashohor. Review of brain mri image segmentation methods. *Artificial Intelligence Review*, 33(3):261–274, 2010.

[BSF94] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

[BXL+18] Xia-an Bi, Qian Xu, Xianhao Luo, Qi Sun, and Zhigang Wang. Analysis of progression toward alzheimer's disease based on evolutionary weighted random support vector machine cluster. *Frontiers in neuroscience*, 12:716, 2018.

[BY11] Deborah E Barnes and Kristine Yaffe. The projected effect of risk factor reduction on alzheimer's disease prevalence. *The Lancet Neurology*, 10(9):819–828, 2011.

[CAD+05] Owen T Carmichael, Howard A Aizenstein, Simon W Davis, James T Becker, Paul M Thompson, Carolyn Cidis Meltzer, and Yanxi Liu. Atlas-based hippocampus segmentation in alzheimer's disease and mild cognitive impairment. *Neuroimage*, 27(4):979–990, 2005.

[CG16] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[CGT+11a] Rémi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehéricy, Marie-Odile Habert, Marie Chupin, Habib Benali, Olivier Colliot, Alzheimer's Disease Neuroimaging Initiative, et al. Automatic classification of patients with alzheimer's disease from structural mri: a comparison of ten methods using the adni database. *neuroimage*, 56(2):766–781, 2011.

[CGT+11b] Rémi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehéricy, Marie-Odile Habert, Marie Chupin, Habib Benali, Olivier Colliot, Alzheimer's Disease Neuroimaging Initiative, et al. Automatic classification of patients with alzheimer's disease from structural mri: a comparison of ten methods using the adni database. *neuroimage*, 56(2):766–781, 2011.

[CL18] Ruoxuan Cui and Manhua Liu. Hippocampus analysis by combination of 3-d densenet and shapes for alzheimer's disease diagnosis. *IEEE journal of biomedical and health informatics*, 23(5):2099–2107, 2018.

[CLH18] Joseph Paul Cohen, Margaux Luck, and Sina Honari. Distribution matching losses can hallucinate features in medical image translation. In *International conference on medical image computing and computer-assisted intervention*, pages 529–536. Springer, 2018.

[CLL18] Ruoxuan Cui, Manhua Liu, and Gang Li. Longitudinal analysis for alzheimer's disease diagnosis using rnn. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1398–1401. IEEE, 2018.

[CLY+21] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[CLZ+18] Liang Cao, Long Li, Jifeng Zheng, Xin Fan, Feng Yin, Hui Shen, and Jun Zhang. Multi-task neural networks for joint hippocampus segmentation and clinical score regression. *Multimedia Tools and Applications*, 77(22):29669–29686, 2018.

[CPC+18] Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):1–12, 2018.

[CPV+15] Sergio Campos, Luis Pizarro, Carlos Valle, Katherine R Gray, Daniel Rueckert, and Héctor Allende. Evaluating imputation techniques for missing data in adni: a patient classification study. In *Iberoamerican Congress on Pattern Recognition*, pages 3–10. Springer, 2015.

[CRBFS19] Daniel E Cahall, Ghulam Rasool, Nidhal C Bouaynaya, and Hassan M Fathallah-Shaykh. Inception modules enhance brain tumor segmentation. *Frontiers in computational neuroscience*, 13:44, 2019.

[CRG+09] R Chaves, J Ramírez, JM Górriz, M López, D Salas-Gonzalez, I Alvarez, and F Segovia. Svm-based computer-aided diagnosis of the alzheimer's disease using t-test nmse feature selection with feature correlation weighting. *Neuroscience letters*, 461(3):293–297, 2009.

[CVMG+14a] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[CVMG+14b] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[DAD18] Jose Dolz, Ismail Ben Ayed, and Christian Desrosiers. Dense multi-path u-net for ischemic stroke lesion segmentation in multiple image modalities. In *International MICCAI Brainlesion Workshop*, pages 271–282. Springer, 2018.

[DBK+20] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[DBS+11] Christos Davatzikos, Priyanka Bhatt, Leslie M Shaw, Kayhan N Batmanghelich, and John Q Trojanowski. Prediction of mci to ad conversion, via mri, csf biomarkers, and pattern classification. *Neurobiology of aging*, 32(12):2322–e19, 2011.

[DÇ20] İbrahim Delibaşoğlu and Müfit Çetin. Building segmentation with inception-unet and classical methods. In *2020 28th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE, 2020.

[DCG+09] Simon Duchesne, Anna Caroli, Cristina Geroldi, D Louis Collins, and Giovanni B Frisoni. Relating one-year cognitive change in mild cognitive impairment to baseline mri features. *Neuroimage*, 47(4):1363–1370, 2009.

[DFA+09] Bradford C Dickerson, Eric Feczko, Jean C Augustinack, Jenni Pacheco, John C Morris, Bruce Fischl, and Randy L Buckner. Differential effects of aging and alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiology of aging*, 30(3):432–440, 2009.

[DHS11] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[DSL+02] AT Du, N Schuff, MP Laakso, XP Zhu, WJ Jagust, K Yaffe, JH Kramer, BL Miller, Bruce R Reed, D Norman, et al. Effects of subcortical ischemic vascular dementia and ad on entorhinal cortex and hippocampus. *Neurology*, 58(11):1635–1641, 2002.

[EBP+18] Rannveig Sakshaug Eldholm, Maria Lage Barca, Karin Persson, Anne-Brita Knapskog, Hege Kersten, Knut Engedal, Geir Selbaek, Anne Braekhus, Eva Skovlund, and Ingvild Saltvedt. Progression of alzheimer's disease: a longitudinal study in norwegian memory clinics. *Journal of Alzheimer's Disease*, 61(3):1221–1232, 2018.

[ETA+20] Mohammad Eslami, Solale Tabarestani, Shadi Albarqouni, Ehsan Adeli, Nassir Navab, and Malek Adjouadi. Image-to-images translation for multi-task organ segmentation and bone suppression in chest x-ray radiography. *IEEE transactions on medical imaging*, 39(7):2553–2565, 2020.

[Fis12] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[FLC+17] Chen Fang, Chunfei Li, Mercedes Cabrerizo, Armando Barreto, Jean Andrian, David Loewenstein, Ranjan Duara, and Malek Adjouadi. A novel gaussian discriminant analysis-based computer aided diagnosis system for screening different stages of alzheimer's disease. In *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 279–284. IEEE, 2017.

[FSB+02] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

[FWLW20] Tongle Fan, Guanglei Wang, Yan Li, and Hongrui Wang. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8:179656–179665, 2020.

[GAH+13a] Katherine R Gray, Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative, et al. Random forest-based similarity measures for multi-modal classification of alzheimer's disease. *NeuroImage*, 65:167–175, 2013.

[GAH+13b] Katherine R Gray, Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative, et al. Random forest-based similarity measures for multi-modal classification of alzheimer's disease. *NeuroImage*, 65:167–175, 2013.

[GCC+09] Emilie Gerardin, Gaël Chételat, Marie Chupin, Rémi Cuingnet, Béatrice Desgranges, Ho-Sung Kim, Marc Niethammer, Bruno Dubois, Stéphane Lehéricy, Line Garnero, et al. Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging. *Neuroimage*, 47(4):1476–1486, 2009.

[GGB+14] Mohammed Goryawala, Seza Gulec, Ruchir Bhatt, Anthony J McGoron, and Malek Adjouadi. A low-interaction automatic 3d liver segmentation method using computed tomography for selective internal radiation therapy. *BioMed research international*, 2014, 2014.

[GGH+18] Eli Gibson, Francesco Giganti, Yipeng Hu, Ester Bonmati, Steve Bandula, Kurinchi Gurusamy, Brian Davidson, Stephen P Pereira, Matthew J Clarkson, and Dean C Barratt. Automatic multi-organ segmentation on abdominal ct with dense v-networks. *IEEE transactions on medical imaging*, 37(8):1822–1834, 2018.

[GGS15] Yanrong Guo, Yaozong Gao, and Dinggang Shen. Deformable mr prostate segmentation via deep feature learning and sparse patch matching. *IEEE transactions on medical imaging*, 35(4):1077–1089, 2015.

[GH15] HT Gorji and J Haddadnia. A novel method for early diagnosis of alzheimer's disease based on pseudo zernike moment from structural mri. *Neuroscience*, 305:361–371, 2015.

[GKH+90] Douglas Galasko, Melville R Klauber, C Richard Hofstetter, David P Salmon, Bruce Lasker, and Leon J Thal. The mini-mental state examination in the early diagnosis of alzheimer's disease. *Archives of neurology*, 47(1):49–52, 1990.

[GR71] Gene H Golub and Christian Reinsch. Singular value decomposition and least squares solutions. In *Linear Algebra*, pages 134–151. Springer, 1971.

[GSRM04] Michael D Greicius, Gaurav Srivastava, Allan L Reiss, and Vinod Menon. Default-mode network activity distinguishes alzheimer's disease from healthy aging: evidence from functional mri. *Proceedings of the National Academy of Sciences*, 101(13):4637–4642, 2004.

[GVOV+16] Sandra González-Villà, Arnau Oliver, Sergi Valverde, Liping Wang, Reyer Zwiggelaar, and Xavier Lladó. A review on brain structures segmentation in magnetic resonance imaging. *Artificial intelligence in medicine*, 73:45–69, 2016.

[GW84] George G Glenner and Caine W Wong. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochemical and biophysical research communications*, 120(3):885–890, 1984.

[HA17] Maryam Hajiesmaeili and Majid Amirfakhrian. A new approach to locate the hippocampus nest in brain mr images. In *2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA)*, pages 140–145. IEEE, 2017.

[HAKEB16] Ehsan Hosseini-Asl, Robert Keynton, and Ayman El-Baz. Alzheimer's disease diagnostics by adaptation of 3d convolutional network. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 126–130. IEEE, 2016.

[HCH+17] Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, and Philip HS Torr. Deeply supervised salient object detection with short connections. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3203–3212, 2017.

[HDD+94] J-J Hauw, SE Daniel, D Dickson, DS Horoupian, K Jellinger, PL Lantos, A McKee, M Tabaton, and I Litvan. Preliminary ninds neuropathologic criteria for steele-richardson-olszewski syndrome (progressive supranuclear palsy). *Neurology*, 44(11):2015–2015, 1994.

[HJG+16a] Lei Huang, Yan Jin, Yaozong Gao, Kim-Han Thung, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Longitudinal clinical score prediction in alzheimer's disease with soft-split sparse regression based random forest. *Neurobiology of aging*, 46:180–191, 2016.

[HJG+16b] Lei Huang, Yan Jin, Yaozong Gao, Kim-Han Thung, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Longitudinal clinical score prediction in alzheimer's disease with soft-split sparse regression based random forest. *Neurobiology of aging*, 46:180–191, 2016.

[HJG+16c] Lei Huang, Yan Jin, Yaozong Gao, Kim-Han Thung, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Longitudinal clinical score prediction in alzheimer's disease with soft-split sparse regression based random forest. *Neurobiology of aging*, 46:180–191, 2016.

[HLVDMW17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[HMC+00] Robert E Hogan, Kevin E Mark, Indrajit Choudhuri, Lei Wang, Sarang Joshi, Michael I Miller, and Richard D Bucholz. Magnetic resonance imaging deformation-based segmentation of the hippocampus in patients with mesial temporal sclerosis and temporal lobe epilepsy. *Journal of digital imaging*, 13(1):217–218, 2000.

[HS97] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[HSB+21] Niels Hansen, Aditya Singh, Claudia Bartels, Frederic Brosseron, Katharina Buerger, Arda C Cetindag, Laura Dobisch, Peter Dechent, Birgit B Ertl-Wagner, Klaus Fliessbach, et al. Hippocampal and hippocampal-subfield volumes from early-onset major depression and bipolar disorder to cognitive decline. *Frontiers in aging neuroscience*, 13:153, 2021.

[HSS18] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[HWB+04] R Edward Hogan, Lei Wang, Mary E Bertrand, L James Willmore, Richard D Bucholz, A Sami Nassif, and John G Csernansky. Mri-based high-dimensional hippocampal mapping in mesial temporal lobe epilepsy. *Brain*, 127(8):1731–1740, 2004.

[HYJW08] Chengzhong Huang, Bin Yan, Hua Jiang, and Dahui Wang. Combining voxel-based morphometry with artifical neural network theory in the application research of diagnosing alzheimer's disease. In *2008 International Conference on BioMedical Engineering and Informatics*, volume 1, pages 250–254. IEEE, 2008.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[IMC+17] W Izquierdo, H Martin, M Cabrerizo, A Barreto, J Andrian, N Rishe, S Gonzalez-Arias, David Loewenstein, R Duara, and M Adjouadi. Robust prediction of cognitive test scores in alzheimer's patients. In *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–7. IEEE, 2017.

[Jag13] William Jagust. Vulnerable neural systems and the borderland of brain aging and neurodegeneration. *Neuron*, 77(2):219–234, 2013.

[Jag18] William Jagust. Imaging the evolution and pathophysiology of alzheimer disease. *Nature Reviews Neuroscience*, 19(11):687–700, 2018.

[JJAK+11] Clifford R Jack Jr, Marilyn S Albert, David S Knopman, Guy M McKhann, Reisa A Sperling, Maria C Carrillo, Bill Thies, and Creighton H Phelps. Introduction to the recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):257–262, 2011.

[JJWS+18] Clifford R Jack Jr, Heather J Wiste, Christopher G Schwarz, Val J Lowe, Matthew L Senjem, Prashanthi Vemuri, Stephen D Weigand, Terry M Therneau, Dave S Knopman, Jeffrey L Gunter, et al. Longitudinal tau pet in ageing and alzheimer's disease. *Brain*, 141(5):1517–1528, 2018.

[JMC+16] Bo Jiang, Shiqian Ma, Jason Causey, Linbo Qiao, Matthew Price Hardin, Ian Bitts, Daniel Johnson, Shuzhong Zhang, and Xiuzhen Huang. Sparrec: An effective matrix completion framework of missing data imputation for gwas. *Scientific reports*, 6:35534, 2016.

[JPM18] T Nimmy John, Subha D Puthankattil, and Ramshekhar Menon. Analysis of long range dependence in the eeg signals of alzheimer patients. *Cognitive neurodynamics*, 12(2):183–199, 2018.

[JPOT92] Clifford R Jack, Ronald C Petersen, Peter C O'brien, and Eric G Tangalos. Mr-based hippocampal volumetry in the diagnosis of alzheimer's disease. *Neurology*, 42(1):183–183, 1992.

[JSR+19] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Dag Johansen, Thomas De Lange, Pål Halvorsen, and Håvard D Johansen. Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*, pages 225–2255. IEEE, 2019.

[JVW+11] Clifford R Jack, Prashanthi Vemuri, Heather J Wiste, Stephen D Weigand, Paul S Aisen, John Q Trojanowski, Leslie M Shaw, Matthew A Bernstein, Ronald C Petersen, Michael W Weiner, et al. Evidence for ordering of alzheimer disease biomarkers. *Archives of neurology*, 68(12):1526–1535, 2011.

[JVW+14] Piers Johnson, Luke Vandewater, William Wilson, Paul Maruff, Greg Savage, Petra Graham, Lance S Macaulay, Kathryn A Ellis, Cassandra Szoeke, Ralph N Martins, et al. Genetic algorithm with logistic regression for prediction of progression to alzheimer's disease. *BMC bioinformatics*, 15(S16):S11, 2014.

[KB14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[KSBD17] Sergey Korolev, Amir Safiullin, Mikhail Belyaev, and Yulia Dodonova. Residual and plain convolutional neural networks for 3d brain mri classification. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 835–838. IEEE, 2017.

[KSH17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[KYP+18] Kichang Kwak, Hyuk Jin Yun, Gilsoon Park, Jong-Min Lee, Alzheimer's Disease Neuroimaging Initiative, et al. Multi-modality sparse representation for alzheimer's disease classification. *Journal of Alzheimer's Disease*, 65(3):807–817, 2018.

[LB14] Salim Lahmiri and Mounir Boukadoum. New approach for automatic classification of alzheimer's disease, mild cognitive impairment and

healthy brain magnetic resonance images. *Healthcare technology letters*, 1(1):32–36, 2014.

[LCD+18] David A Loewenstein, Rosie E Curiel, Steven DeKosky, Russell M Bauer, Monica Rosselli, Salvador M Guinjoan, Malek Adjouadi, Ailyn Peñate, William W Barker, Sindy Goenaga, et al. Utilizing semantic intrusions to identify amyloid positivity in mild cognitive impairment. *Neurology*, 91(10):e976–e984, 2018.

[LCQ+18] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE transactions on medical imaging*, 37(12):2663–2674, 2018.

[LGH12] Julián Luengo, Salvador García, and Francisco Herrera. On the choice of the best imputation methods for missing values considering three groups of classification methods. *Knowledge and information systems*, 32(1):77–108, 2012.

[LHM+10] SM Landau, D Harvey, CM Madison, EM Reiman, NL Foster, PS Aisen, Ronald Carl Petersen, LM Shaw, JQ Trojanowski, CR Jack, et al. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*, 75(3):230–238, 2010.

[LHW+19] Hongming Li, Mohamad Habes, David A Wolk, Yong Fan, Alzheimer's Disease Neuroimaging Initiative, et al. A deep learning model for early prediction of alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's & Dementia*, 15(8):1059–1070, 2019.

[LJ12] Raymond Y Lo and William J Jagust. Predicting missing biomarker data in a longitudinal study of alzheimer disease. *Neurology*, 78(18):1376–1382, 2012.

[LLC+14] Siqi Liu, Sidong Liu, Weidong Cai, Hangyu Che, Sonia Pujol, Ron Kikinis, Dagan Feng, Michael J Fulham, et al. Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease. *IEEE Transactions on Biomedical Engineering*, 62(4):1132–1140, 2014.

[LLI+19] Fan Li, Manhua Liu, Alzheimer's Disease Neuroimaging Initiative, et al. A hybrid convolutional and recurrent neural network for hip-

pocampus analysis in alzheimer's disease. *Journal of neuroscience methods*, 323:108–118, 2019.

[LLY+20] Manhua Liu, Fan Li, Hao Yan, Kundong Wang, Yixin Ma, Li Shen, Mingqing Xu, Alzheimer's Disease Neuroimaging Initiative, et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in alzheimer's disease. *NeuroImage*, 208:116459, 2020.

[LLZS18] Chunfeng Lian, Mingxia Liu, Jun Zhang, and Dinggang Shen. Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural mri. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):880–893, 2018.

[LMW+21] Yu Liu, Jie Meng, Kangcheng Wang, Kaixiang Zhuang, Qunlin Chen, Wenjing Yang, Jiang Qiu, and Dongtao Wei. Morphometry of the hippocampus across the adult life-span in patients with depressive disorders: Association with neuroticism. *Brain Topography*, pages 1–11, 2021.

[LO17] Yin Lou and Mikhail Obukhov. Bdt: Gradient boosted decision tables for high accuracy and scoring efficiency. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1893–1901, 2017.

[LSD15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[LW+02] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[LWVW+14] AV Lebedev, Eric Westman, GJP Van Westen, MG Kramberger, Arvid Lundervold, Dag Aarsland, H Soininen, I Kłoszewska, P Mecocci, M Tsolaki, et al. Random forest ensembles for detection and prediction of alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clinical*, 6:115–125, 2014.

[LWW+12] Yang Li, Yaping Wang, Guorong Wu, Feng Shi, Luping Zhou, Weili Lin, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative,

et al. Discriminant analysis of longitudinal cortical thickness changes in alzheimer's disease using dynamic and network features. *Neurobiology of aging*, 33(2):427–e15, 2012.

[LXC+17] Shen Lu, Yong Xia, Weidong Cai, Michael Fulham, David Dagan Feng, Alzheimer's Disease Neuroimaging Initiative, et al. Early identification of mild cognitive impairment using incomplete random forest-robust support vector machine and fdg-pet imaging. *Computerized Medical Imaging and Graphics*, 60:35–41, 2017.

[LZAS17] Mingxia Liu, Jun Zhang, Ehsan Adeli, and Dinggang Shen. Deep multi-task multi-channel learning for joint classification and regression of brain status. In *International conference on medical image computing and computer-assisted intervention*, pages 3–11. Springer, 2017.

[LZAS18] Mingxia Liu, Jun Zhang, Ehsan Adeli, and Dinggang Shen. Landmark-based deep multi-instance learning for brain disease diagnosis. *Medical image analysis*, 43:157–168, 2018.

[MCL+15a] Francesco Carlo Morabito, Maurizio Campolo, Domenico Labate, Giuseppe Morabito, Lilla Bonanno, Alessia Bramanti, Simona De Salvo, Angela Marra, and Placido Bramanti. A longitudinal eeg study of alzheimer's disease progression based on a complex network approach. *International journal of neural systems*, 25(02):1550005, 2015.

[MCL+15b] Francesco Carlo Morabito, Maurizio Campolo, Domenico Labate, Giuseppe Morabito, Lilla Bonanno, Alessia Bramanti, Simona De Salvo, Angela Marra, and Placido Bramanti. A longitudinal eeg study of alzheimer's disease progression based on a complex network approach. *International journal of neural systems*, 25(02):1550005, 2015.

[MHH+17] Elaheh Moradi, Ilona Hallikainen, Tuomo Hänninen, Jussi Tohka, Alzheimer's Disease Neuroimaging Initiative, et al. Rey's auditory verbal learning test scores can be predicted from whole brain mri in alzheimer's disease. *NeuroImage: Clinical*, 13:415–427, 2017.

[MHT10] Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research*, 11:2287–2322, 2010.

[MID+19] Niklas Mattsson, Philip S Insel, Michael Donohue, Jonas Jögi, Rik Ossenkoppele, Tomas Olsson, Michael Schöll, Ruben Smith, and Oskar Hansson. Predicting diagnosis and cognition with 18f-av-1451 tau pet and structural mri in alzheimer's disease. *Alzheimer's & Dementia*, 15(4):570–580, 2019.

[MMC+19] Mehdi Mafi, Harold Martin, Mercedes Cabrerizo, Jean Andrian, Armando Barreto, and Malek Adjouadi. A comprehensive survey on impulse and gaussian denoising filters for digital images. *Signal Processing*, 157:236–260, 2019.

[MMK+09] Benoît Magnin, Lilia Mesrob, Serge Kinkingnéhun, Mélanie Pélégrini-Issac, Olivier Colliot, Marie Sarazin, Bruno Dubois, Stéphane Lehéricy, and Habib Benali. Support vector machine-based classification of alzheimer's disease from whole-brain anatomical mri. *Neuroradiology*, 51(2):73–83, 2009.

[MPG+15a] Elaheh Moradi, Antonietta Pepe, Christian Gaser, Heikki Huttunen, Jussi Tohka, Alzheimer's Disease Neuroimaging Initiative, et al. Machine learning framework for early mri-based alzheimer's conversion prediction in mci subjects. *Neuroimage*, 104:398–412, 2015.

[MPG+15b] Elaheh Moradi, Antonietta Pepe, Christian Gaser, Heikki Huttunen, Jussi Tohka, Alzheimer's Disease Neuroimaging Initiative, et al. Machine learning framework for early mri-based alzheimer's conversion prediction in mci subjects. *Neuroimage*, 104:398–412, 2015.

[MRCA18] Mehdi Mafi, Hoda Rajaei, Mercedes Cabrerizo, and Malek Adjouadi. A robust edge detection approach in the presence of high impulse noise intensity through switching adaptive median and fixed weighted mean filtering. *IEEE Transactions on Image Processing*, 27(11):5475–5490, 2018.

[NRKL10] Agneta Nordberg, Juha O Rinne, Ahmadul Kadir, and Bengt Långström. The use of pet in alzheimer disease. *Nature Reviews Neurology*, 6(2):78–87, 2010.

[NTL+17] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 417–425. Springer, 2017.

[NZM+16] Liqiang Nie, Luming Zhang, Lei Meng, Xuemeng Song, Xiaojun Chang, and Xuelong Li. Modeling disease progression via multisource multitask learners: A case study with alzheimer's disease. *IEEE transactions on neural networks and learning systems*, 28(7):1508–1519, 2016.

[OJMM19] Caroline Vintergaard Ott, Claire Bergstrom Johnson, Julian Macoveanu, and Kamilla Miskowiak. Structural changes in the hippocampus as a biomarker for cognitive improvements in neuropsychiatric disorders: A systematic review. *European Neuropsychopharmacology*, 29(3):319–329, 2019.

[OMÁI+15] Andrés Ortiz, Jorge Munilla, Ignacio Álvarez-Illán, Juan M Górriz, Javier Ramírez, Alzheimer's Disease Neuroimaging Initiative, et al. Exploratory graphical models of functional and structural connectivity patterns for alzheimer's disease diagnosis. *Frontiers in computational neuroscience*, 9:132, 2015.

[OPSS11] Joseph O Ogutu, Hans-Peter Piepho, and Torben Schulz-Streeck. A comparison of random forests, boosting and support vector machines for genomic selection. In *BMC proceedings*, volume 5, page S11. Springer, 2011.

[PDHvdF+13] Simon-Shlomo Poil, Willem De Haan, Wiesje M van der Flier, Huibert D Mansvelder, Philip Scheltens, and Klaus Linkenkaer-Hansen. Integrative eeg biomarkers predict progression to alzheimer's disease at the mci stage. *Frontiers in aging neuroscience*, 5:58, 2013.

[PM05a] Ronald C Petersen and John C Morris. Mild cognitive impairment as a clinical entity and treatment target. *Archives of neurology*, 62(7):1160–1163, 2005.

[PM05b] Ronald C Petersen and John C Morris. Mild cognitive impairment as a clinical entity and treatment target. *Archives of neurology*, 62(7):1160–1163, 2005.

[PM15] Adrien Payan and Giovanni Montana. Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks. *arXiv preprint arXiv:1502.02506*, 2015.

[PPW+14] Jon Pipitone, Min Tae M Park, Julie Winterburn, Tristram A Lett, Jason P Lerch, Jens C Pruessner, Martin Lepage, Aristotle N

Voineskos, M Mallar Chakravarty, Alzheimer's Disease Neuroimaging Initiative, et al. Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. *Neuroimage*, 101:494–512, 2014.

[PT94] Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994.

[PVG+11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[QJZ+20] Saqib Qamar, Hai Jin, Ran Zheng, Parvez Ahmad, and Mohd Usama. A variant form of 3d-unet for infant brain segmentation. *Future Generation Computer Systems*, 108:613–623, 2020.

[RDS+15] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[RM14] Christiane Reitz and Richard Mayeux. Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochemical pharmacology*, 88(4):640–651, 2014.

[RSW+15] Kerstin Ritter, Julia Schumacher, Martin Weygandt, Ralph Buchert, Carsten Allefeld, John-Dylan Haynes, Alzheimer's Disease Neuroimaging Initiative, et al. Multimodal prediction of conversion to alzheimer's disease based on incomplete biomarkers. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(2):206–215, 2015.

[SA11] Ziad Sankari and Hojjat Adeli. Probabilistic neural networks for diagnosis of alzheimer's disease using conventional and wavelet coherence. *Journal of neuroscience methods*, 197(1):165–170, 2011.

[SAB+11] Reisa A Sperling, Paul S Aisen, Laurel A Beckett, David A Bennett, Suzanne Craft, Anne M Fagan, Takeshi Iwatsubo, Clifford R Jack Jr, Jeffrey Kaye, Thomas J Montine, et al. Toward defining the preclinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):280–292, 2011.

[SAB+19] Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019.

[SCQ17] Alessia Sarica, Antonio Cerasa, and Aldo Quattrone. Random forest algorithm for the classification of neuroimaging data in alzheimer's disease: A systematic review. *Frontiers in aging neuroscience*, 9:329, 2017.

[SCZ+17] Bibo Shi, Yani Chen, Pin Zhang, Charles D Smith, Jundong Liu, Alzheimer's Disease Neuroimaging Initiative, et al. Nonlinear feature transformation and deep fusion for alzheimer's disease staging analysis. *Pattern recognition*, 63:487–498, 2017.

[SGR+10] F Segovia, JM Górriz, J Ramírez, D Salas-González, I Álvarez, M López, R Chaves, and P Padilla. Classification of functional brain images using a gmm-based multi-variate approach. *Neuroscience Letters*, 474(1):58–62, 2010.

[SLPG04] Martin Styner, Jeffrey A Lieberman, Dimitrios Pantazis, and Guido Gerig. Boundary and medial shape analysis of the hippocampus in schizophrenia. *Medical image analysis*, 8(3):197–203, 2004.

[SLS+14a] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. *NeuroImage*, 101:569–582, 2014.

[SLS+14b] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. *NeuroImage*, 101:569–582, 2014.

[SLS+16] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Deep sparse multi-task learning for feature selection in alzheimer's disease diagnosis. *Brain Structure and Function*, 221(5):2569–2587, 2016.

[SLV+17] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 240–248. Springer, 2017.

[SS13] Heung-Il Suk and Dinggang Shen. Deep learning-based feature representation for ad/mci classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 583–590. Springer, 2013.

[STC+18] Danail Stoyanov, Zeike Taylor, Gustavo Carneiro, Tanveer Syeda-Mahmood, Anne Martel, Lena Maier-Hein, João Manuel RS Tavares, Andrew Bradley, João Paulo Papa, Vasileios Belagiannis, et al. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*, volume 11045. Springer, 2018.

[SV99] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[SZL+17] Jun Shi, Xiao Zheng, Yan Li, Qi Zhang, and Shihui Ying. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimer's disease. *IEEE journal of biomedical and health informatics*, 22(1):173–183, 2017.

[TAE+20] Solale Tabarestani, Maryamossadat Aghili, Mohammad Eslami, Mercedes Cabrerizo, Armando Barreto, Naphtali Rishe, Rosie E Curiel, David Loewenstein, Ranjan Duara, and Malek Adjouadi.

A distributed multitask multimodal approach for the prediction of alzheimer's disease in a longitudinal study. *NeuroImage*, 206:116317, 2020.

[TAS+18] Solale Tabarestani, Maryamossadat Aghili, Mehdi Shojaie, Christian Freytes, and Malek Adjouadi. Profile-specific regression model for progression prediction of alzheimer's disease using longitudinal data. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1353–1357. IEEE, 2018.

[TAS+19] Solale Tabarestani, Maryamossadat Aghili, Mehdi Shojaie, Christian Freytes, Mercedes Cabrerizo, Armando Barreto, Naphtali Rishe, Rosie E Curiel, David Loewenstein, Ranjan Duara, et al. Longitudinal prediction modeling of alzheimer disease using recurrent neural networks. In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 1–4. IEEE, 2019.

[TCS+01] Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B Altman. Missing value estimation methods for dna microarrays. *Bioinformatics*, 17(6):520–525, 2001.

[TH+12] Tijmen Tieleman, Geoffrey Hinton, et al. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.

[TL19] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

[TLZJ17] Luan Tran, Xiaoming Liu, Jiayu Zhou, and Rong Jin. Missing modalities imputation via cascaded residual autoencoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1405–1414, 2017.

[TRK+20] M Tanveer, B Richhariya, RU Khan, AH Rashid, P Khanna, M Prasad, and CT Lin. Machine learning techniques for the diagnosis of alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(1s):1–35, 2020.

[TYKM05] Mary C Tierney, Christie Yao, Alex Kiss, and Ian McDowell. Neuropsychological tests accurately predict incident alzheimer disease after 5 and 10 years. *Neurology*, 64(11):1853–1859, 2005.

[vdLDHBN08] Fedde van der Lijn, Tom Den Heijer, Monique MB Breteler, and Wiro J Niessen. Hippocampus segmentation in mr images using atlas registration, voxel classification, and graph cuts. *Neuroimage*, 43(4):708–720, 2008.

[VGS+08] Prashanthi Vemuri, Jeffrey L Gunter, Matthew L Senjem, Jennifer L Whitwell, Kejal Kantarci, David S Knopman, Bradley F Boeve, Ronald C Petersen, and Clifford R Jack Jr. Alzheimer's disease diagnosis in individual subjects using structural mr images: validation studies. *Neuroimage*, 39(3):1186–1197, 2008.

[WBP+17] Dominik Wolf, Martina Bocchetta, Gregory M Preboske, Marina Boccardi, Michel J Grothe, Alzheimer's Disease Neuroimaging Initiative, et al. Reference standard space hippocampus labels according to the european alzheimer's disease consortium–alzheimer's disease neuroimaging initiative harmonized protocol: Utility in automated volumetry. *Alzheimer's & Dementia*, 13(8):893–902, 2017.

[WCY+16] Pingyue Wang, Kewei Chen, Li Yao, Bin Hu, Xia Wu, Jiacai Zhang, Qing Ye, and Xiaojuan Guo. Multimodal classification of mild cognitive impairment based on partial least squares. *Journal of Alzheimer's Disease*, 54(1):359–371, 2016.

[WDD+12] Daniel Weintraub, Nicole Dietz, John E Duda, David A Wolk, Jimit Doshi, Sharon X Xie, Christos Davatzikos, Christopher M Clark, and Andrew Siderowf. Alzheimer's disease pattern of brain atrophy predicts cognitive decline in parkinson's disease. *Brain*, 135(1):170–180, 2012.

[WNK+21] Davis C Woodworth, Hannah L Nguyen, Zainab Khan, Claudia H Kawas, María M Corrada, and S Ahmad Sajjadi. Utility of mri in the identification of hippocampal sclerosis of aging. *Alzheimer's & Dementia*, 17(5):847–855, 2021.

[WT21] Jiong Wu and Xiaoying Tang. Brain segmentation based on multi-atlas and diffeomorphism guided 3d fully convolutional network ensembles. *Pattern Recognition*, 115:107904, 2021.

[WWS+07] Jennifer L Whitwell, Stephen D Weigand, Maria M Shiung, Bradley F Boeve, Tanis J Ferman, Glenn E Smith, David S Knopman, Ronald C Petersen, Eduardo E Benarroch, Keith A Josephs, et al. Focal atrophy in dementia with lewy bodies on mri: a distinct pattern from alzheimer's disease. *Brain*, 130(3):708–719, 2007.

[WZL+18] Xiaoqian Wang, Xiantong Zhen, Quanzheng Li, Dinggang Shen, and Heng Huang. Cognitive assessment prediction in alzheimer's disease by multi-layer multi-target regression. *Neuroinformatics*, 16(3-4):285–294, 2018.

[XYF+13] Shuo Xiang, Lei Yuan, Wei Fan, Yalin Wang, Paul M Thompson, and Jieping Ye. Multi-source learning with block-wise missing data for alzheimer's disease prediction. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 185–193, 2013.

[Yak20] Pavel Yakubovskiy. Segmentation models pytorch. *GitHub Repos*, 2020.

[YSSR21] Michael Yeung, Evis Sala, Carola-Bibiane Schönlieb, and Leonardo Rundo. A mixed focal loss function for handling class imbalanced medical image segmentation. *arXiv preprint arXiv:2102.04525*, 2021.

[YXW+18] Qihang Yu, Lingxi Xie, Yan Wang, Yuyin Zhou, Elliot K Fishman, and Alan L Yuille. Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8280–8289, 2018.

[ZBL+20] P Zach, A Bartoš, A Lagutina, Z Wurst, P Gallina, T Rai, K Kieslich, J Riedlová, I Ibrahim, J Tintěra, et al. Easy identification of optimal coronal slice on brain magnetic resonance imaging to measure hippocampal area in alzheimer's disease patients. *BioMed Research International*, 2020, 2020.

[ZCN+17] Qiao Zhang, Zhipeng Cui, Xiaoguang Niu, Shijie Geng, and Yu Qiao. Image segmentation with pyramid dilated convolution based on resnet and u-net. In *International Conference on Neural Information Processing*, pages 364–372. Springer, 2017.

[ZGG+16] Jun Zhang, Yue Gao, Yaozong Gao, Brent C Munsell, and Dinggang Shen. Detecting anatomical landmarks for fast alzheimer's disease diagnosis. *IEEE transactions on medical imaging*, 35(12):2524–2533, 2016.

[Zha12] Shichao Zhang. Nearest neighbor selection for iteratively knn imputation. *Journal of Systems and Software*, 85(11):2541–2552, 2012.

[ZLA+17] Jun Zhang, Mingxia Liu, Le An, Yaozong Gao, and Dinggang Shen. Alzheimer's disease diagnosis using landmark-based features from longitudinal structural mr images. *IEEE journal of biomedical and health informatics*, 21(6):1607–1616, 2017.

[ZSI+12a] Daoqiang Zhang, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. *NeuroImage*, 59(2):895–907, 2012.

[ZSI+12b] Daoqiang Zhang, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. *NeuroImage*, 59(2):895–907, 2012.

[ZSTL18] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.

[ZSW+17] Xiaofeng Zhu, Heung-Il Suk, Li Wang, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. A novel relational regularization feature selection method for joint regression and classification in ad diagnosis. *Medical image analysis*, 38:205–214, 2017.

[ZTA+17] Xiaofeng Zhu, Kim-Han Thung, Ehsan Adeli, Yu Zhang, and Dinggang Shen. Maximum mean discrepancy based multiple kernel learning for incomplete multimodality neuroimaging data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 72–80. Springer, 2017.

[ZWZ+11] Daoqiang Zhang, Yaping Wang, Luping Zhou, Hong Yuan, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Multimodal classification of alzheimer's disease and mild cognitive impairment. *Neuroimage*, 55(3):856–867, 2011.

[ZXS+17] Yuyin Zhou, Lingxi Xie, Wei Shen, Yan Wang, Elliot K Fishman, and Alan L Yuille. A fixed-point model for pancreas segmentation in abdominal ct scans. In *International conference on medical image computing and computer-assisted intervention*, pages 693–701. Springer, 2017.

[ZYLY11a] Jiayu Zhou, Lei Yuan, Jun Liu, and Jieping Ye. A multi-task learning formulation for predicting disease progression. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 814–822, 2011.

[ZYLY11b] Jiayu Zhou, Lei Yuan, Jun Liu, and Jieping Ye. A multi-task learning formulation for predicting disease progression. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 814–822, 2011.

[ZZ19] Teng Zhang and Zhi-Hua Zhou. Optimal margin distribution machine. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1143–1156, 2019.

Supplementary Documentation

Volumetric Calculations of the Hippocampus Region using the Mount Sinai Medical Center (MSMC) Data as Part of the 1Florida Alzheimer's Disease Research Center (ADRC)

Difference ((Left) Difference ((Right) Difference ((Left+Right)

Mean 6.90 6.57 6.73

STD 5.58 4.33 4.96

#	PID	Date	model(L)	model(R)	FS(L)	FS(R)	Diff(L)	Diff(R)
1	004	02/2017	3189.4	3771.8	3260.6	3800.9	0.021834	0.007725
2	004	06/2009	3401.5	3906.2	3521.2	3989.4	0.033994	0.021308
3	005	02/2017	3867.1	4359.3	4040.1	4087.9	0.042829	0.062250
4	008	03/2006	3187.4	3420.5	3466.8	3438.9	0.080597	0.005379
5	015	02/2017	3146.6	4083.8	3494.1	3602.4	0.099461	0.117888
6	016	02/2017	2839.2	3352.9	2992.2	3611.1	0.051127	0.076998
7	020	03/2018	3585.9	4398.1	3928.1	4198.8	0.087105	0.045321
8	025	04/2016	3841.6	3589.4	3589.6	3631.9	0.070209	0.011849
9	026	04/2016	4045.1	4195.1	4307.1	3919.8	0.060819	0.065621
10	031	04/2017	2880.2	3626.6	3456.3	3503.3	0.166676	0.033996
11	033	04/2016	3677.8	3371.6	3846.1	3761.8	0.043748	0.115740
12	034	04/2016	2463.1	2496.0	2749.4	2321.9	0.104126	0.069734
13	036	07/2017	4370.7	4176.9	4334.8	4493.4	0.008281	0.075781
14	037	08/2016	3506.0	3496.8	3493.1	3792.1	0.003687	0.084447
15	038	06/2016	3414.5	4309.0	3323.8	4161.3	0.027303	0.034271
16	041	04/2016	2820.0	3884.8	3414.5	3611.3	0.174097	0.070399
17	042	05/2016	3515.4	3162.5	3552.2	3534.2	0.010365	0.117525
18	044	05/2016	3769.0	4188.3	3705.5	3577.4	0.017146	0.145867
19	045	05/2016	3183.2	4218.3	3447.6	3597.3	0.076690	0.147220
20	046	04/2016	3742.7	3231.0	3705.5	3577.4	0.010031	0.107219
21	047	05/2016	3418.1	3783.4	3489.2	3768.6	0.020369	0.003921
22	048	05/2016	3748.1	3700.8	3646.5	3760.7	0.027875	0.016180
23	049	04/2016	3196.6	3790.1	3526.0	3656.5	0.093428	0.035258
24	050	05/2017	2896.7	4068.1	3411.1	3624.2	0.150797	0.109113
25	051	03/2016	3003.6	1600.7	3228.7	1424.5	0.069715	0.110074
26	052	05/2016	3071.9	1978.5	3063.5	1762.1	0.002746	0.109370
27	053	10/2016	2585.6	2799.3	2596.4	2976.9	0.004168	0.063435
28	054	06/2016	3185.2	2009.4	3117.7	1854.1	0.021638	0.077272
29	055	04/2016	2913.1	3811.9	3440.6	3529.5	0.153326	0.074088
30	057	05/2016	3554.8	3804.4	3509.9	3523.4	0.012803	0.073869

#	PID	Date	model(L)	model(R)	FS(L)	FS(R)	Diff(L)	Diff(R)
31	059	04/2016	3280.4	3452.1	3290.4	3482.7	0.003045	0.008850
32	060	05/2016	2808.8	3921.3	3183.5	3333.4	0.117710	0.149919
33	061	05/2016	3959.2	4006.0	3974.0	4237.4	0.003724	0.057764
34	062	04/2016	3763.9	3990.6	3748.8	4076.7	0.004035	0.021581
35	063	04/2016	3612.8	4079.0	3689.4	4075.4	0.020771	0.000879
36	064	07/2016	3564.8	3482.6	3168.5	3553.4	0.125071	0.020342
37	065	04/2016	3748.9	3642.5	3768.1	4046.2	0.005094	0.110845
38	066	06/2016	4179.9	4178.7	4135.2	4349.9	0.010803	0.040971
39	068	04/2016	3660.5	4403.7	3907.4	4358.7	0.063181	0.010212
40	069	07/2016	3080.1	3362.3	3186.3	3397.9	0.033345	0.010595
41	070	05/2016	3374.1	3564.2	3334.7	3500.2	0.011808	0.017962
42	071	05/2016	3720.3	4075.8	3730.4	3573.0	0.002715	0.123371
43	072	05/2016	3500.3	3219.0	3679.4	3580.4	0.048668	0.112271
44	073	05/2016	3157.7	4148.9	3595.7	3674.9	0.121826	0.114250
45	074	05/2016	3746.1	3561.9	3730.0	3796.9	0.004312	0.065962
46	075	05/2016	3532.8	3665.8	3687.4	3681.3	0.041914	0.004216
47	077	05/2016	3772.9	3959.6	3794.6	3664.2	0.005728	0.074606
48	078	05/2016	3272.5	4075.8	3464.2	3512.2	0.055348	0.138278
49	079	06/2016	3654.5	3118.2	3647.8	3577.5	0.001841	0.147304
50	080	05/2016	3354.5	2991.8	2860.1	3006.4	0.172865	0.004876
51	081	05/2016	3198.6	2758.3	2804.2	2922.3	0.140653	0.059475
52	082	05/2016	3110.4	3000.3	2904.5	3135.4	0.070883	0.045033
53	083	08/2016	2853.2	2960.6	2900.6	3018.5	0.016339	0.019542
54	084	05/2016	3165.2	2699.4	2727.2	2827.0	0.160589	0.047270
55	086	05/2016	3689.7	3295.1	3647.8	3577.5	0.011487	0.085716
56	087	06/2016	3650.0	3787.0	3686.7	3797.7	0.009953	0.002824
57	088	06/2016	3690.9	3557.9	3535.9	3656.2	0.043836	0.027633
58	090	06/2016	4004.8	3656.8	4009.9	3856.0	0.001271	0.054465
59	091	07/2016	3710.9	3648.5	3850.4	3839.3	0.036226	0.052286
60	095	07/2016	4001.5	4007.8	4038.5	4153.2	0.009164	0.036279
61	096	07/2016	2883.3	3225.7	3185.0	2811.1	0.094722	0.128528
62	097	07/2016	3686.9	2891.3	3407.7	3207.4	0.081942	0.109329
63	098	07/2016	3401.5	4239.3	4027.8	3937.8	0.155503	0.071122
64	100	09/2016	3507.8	3294.4	3537.8	3076.6	0.008494	0.066103
65	101	07/2014	3968.6	4537.3	3986.8	4028.7	0.004555	0.112085
66	102	08/2016	3377.1	3216.1	3397.3	3679.0	0.005949	0.143940
67	103	07/2016	3444.2	3673.8	3493.0	3861.3	0.013958	0.051043
68	104	07/2016	3395.8	3246.4	3391.4	3698.4	0.001302	0.139247
69	105	07/2016	3548.1	4064.6	3492.0	3764.5	0.016061	0.073835
70	106	08/2016	3020.8	4054.4	3299.7	3664.9	0.084513	0.096062

#	PID	Date	model(L)	model(R)	FS(L)	FS(R)	Diff(L)	Diff(R)
71	108	07/2016	3981.8	4039.0	3919.3	3935.3	0.015945	0.025669
72	109	08/2016	3031.0	3993.9	3729.2	3899.7	0.187214	0.023590
73	111	07/2016	3355.8	2704.1	3321.7	2944.7	0.010259	0.088991
74	112	08/2016	2857.9	2640.2	3353.1	2849.0	0.147670	0.079094
75	115	08/2016	3565.1	2836.1	3216.6	2952.6	0.108338	0.041061
77	118	09/2016	2921.9	2791.5	3137.8	2987.5	0.068801	0.070228
78	120	08/2016	2811.6	2746.3	2857.8	2865.5	0.016178	0.043387
79	122	09/2016	3295.2	2836.4	3137.2	2915.5	0.050361	0.027884
80	123	09/2016	2681.4	3000.0	3062.2	2989.3	0.124366	0.003568
81	124	09/2016	3843.8	3983.8	3677.3	4112.2	0.045289	0.032223
82	125	09/2016	3891.3	3934.1	3760.4	3946.1	0.034812	0.003053
83	126	09/2016	4261.4	4224.2	4137.0	4361.3	0.030069	0.032458
84	127	09/2016	3466.5	4175.1	4240.4	4408.4	0.182510	0.055884
85	128	09/2016	3890.1	4172.1	3995.5	3876.5	0.026378	0.070850
86	131	09/2016	3205.8	4101.0	3680.5	3799.9	0.128977	0.073421
87	132	09/2016	3556.0	3814.2	3603.1	3731.0	0.013076	0.021810
88	133	11/2016	3400.7	3556.0	3411.8	3703.8	0.003248	0.041564
89	135	10/2016	3590.2	3598.5	3473.0	3642.0	0.033744	0.012090
90	136	10/2016	3180.6	3426.0	3376.2	3421.2	0.057933	0.001404
91	138	09/2016	3395.3	3567.8	3281.0	3475.0	0.034850	0.026014
92	139	10/2016	3952.7	3500.5	3307.9	3461.6	0.194914	0.011116
93	140	01/2017	3398.3	4148.1	3815.0	3973.1	0.109237	0.042178
94	141	11/2016	3643.6	3478.5	3388.0	3463.0	0.075450	0.004462
95	142	10/2016	3236.7	3001.9	3207.5	3270.7	0.009095	0.089545
96	143	10/2016	3463.8	3300.7	3437.8	3449.2	0.007560	0.044997
97	144	10/2016	4170.7	4073.6	3985.1	4141.3	0.046575	0.016615
98	145	10/2016	3319.0	4259.3	3384.2	3731.5	0.019260	0.123913
99	146	10/2016	3591.6	4410.7	3591.8	4110.1	0.000061	0.068156
100	147	12/2016	3424.7	3787.4	3549.2	3917.5	0.035071	0.034347
101	148	10/2016	3281.8	4005.1	3590.6	3571.3	0.085995	0.108321
102	149	11/2016	3390.5	4227.4	3759.7	4068.8	0.098193	0.037517
103	150	11/2016	3808.4	3945.3	3892.6	3832.0	0.021641	0.028708
104	151	10/2016	2865.1	2558.2	2739.4	2511.7	0.045884	0.018196
105	152	10/2016	2870.6	3177.7	2899.0	2948.5	0.009799	0.072122

#	PID	Date	model(L)	model(R)	FS(L)	FS(R)	Diff(L)	Diff(R)
106	153	12/2016	2843.1	2810.7	2822.3	2628.4	0.007358	0.064864
107	154	11/2016	3146.1	3091.4	2845.5	2680.9	0.105628	0.132780
108	155	10/2016	3617.2	3790.0	3295.4	3294.7	0.097659	0.130689
109	157	11/2016	2007.4	2069.9	2098.3	2055.3	0.043335	0.007054
110	158	11/2016	3691.5	3517.0	3556.2	3584.3	0.038049	0.019121
111	160	11/2016	3212.9	3151.3	2717.6	3153.2	0.182241	0.000609
112	161	11/2016	3526.5	3741.7	3628.1	3769.1	0.028014	0.007315
113	163	02/2017	3279.0	3915.6	3116.0	3481.2	0.052304	0.110941
114	164	11/2016	3709.2	2564.1	3415.4	2684.6	0.086023	0.047014
115	165	11/2016	2956.1	2085.3	2553.6	2336.0	0.157615	0.120208
116	166	11/2016	2777.2	2711.1	2623.0	2394.4	0.058802	0.116808
117	167	11/2016	2636.8	2496.6	2606.3	2275.3	0.011719	0.088655
118	169	12/2016	3138.8	3223.8	3170.5	3448.3	0.010005	0.069626
119	170	12/2016	2768.6	3366.2	2897.2	3310.9	0.044403	0.016441
120	171	02/2017	4117.6	4598.3	4573.2	4911.8	0.099629	0.068180
121	172	09/2016	3999.4	3773.9	3968.5	4114.2	0.007781	0.090186
122	173	11/2016	2880.1	3113.5	2988.1	3406.5	0.036139	0.094109
123	174	12/2016	3312.6	3857.4	3523.5	3795.6	0.059860	0.016024
124	175	11/2016	2650.3	2617.3	2892.2	2556.4	0.083623	0.023272
125	176	12/2016	2998.0	3917.9	2980.6	3464.1	0.005851	0.115833
126	177	12/2016	3746.7	4678.0	3987.7	4173.0	0.060429	0.107948
127	178	02/2017	3503.2	2691.1	3382.0	2862.1	0.035832	0.063536
128	179	12/2016	3498.0	3365.3	3492.5	3327.1	0.001581	0.011362
129	181	12/2016	2373.3	3328.9	2860.1	3006.4	0.170201	0.096871
130	182	12/2016	3320.5	3119.3	2804.2	2922.3	0.184129	0.063165
131	183	12/2016	3310.3	3178.6	2904.5	3135.4	0.139717	0.013596
132	184	12/2016	2871.5	3209.1	2900.6	3018.5	0.010040	0.059394
133	185	12/2016	3916.8	4041.3	3910.0	4481.2	0.001736	0.108853
134	186	12/2016	3277.4	4104.2	3647.8	3577.5	0.101535	0.128325
135	187	01/2017	3467.9	4384.0	3686.7	3797.7	0.059362	0.133733
136	188	01/2017	3106.0	4021.9	3535.9	3656.2	0.121572	0.090919
137	189	01/2017	2953.6	3617.9	3535.9	3656.2	0.164669	0.010599
138	190	01/2017	3967.4	3966.8	4009.9	3856.0	0.010602	0.027944
139	191	01/2017	3113.8	3827.6	3850.4	3839.3	0.191304	0.003058
140	192	12/2016	3170.5	3246.1	2770.5	3046.1	0.144380	0.061612
141	194	03/2017	2611.9	3443.8	2770.5	3046.1	0.057247	0.115473
142	195	01/2017	3979.1	3699.6	4038.5	4153.2	0.014710	0.122623

#	PID	Date	model(L)	model(R)	FS(L)	FS(R)	Diff(L)	Diff(R)
143	196	01/2017	2980.9	2832.2	3185.0	2811.1	0.064074	0.007442
144	197	03/2017	3007.2	3100.9	3407.7	3207.4	0.117526	0.034361
145	198	02/2017	4011.6	3976.4	4027.8	3937.8	0.004011	0.009704
146	199	01/2017	2862.6	2802.5	2730.9	2914.8	0.048208	0.040088
147	201	02/2017	2893.1	4239.5	3605.8	3692.9	0.197650	0.128925
148	202	02/2017	3072.3	3490.6	3512.4	3811.6	0.125313	0.091972
149	203	02/2017	2407.0	2700.8	2182.7	2729.1	0.102752	0.010482
150	204	02/2017	2297.1	2593.8	2649.0	2624.4	0.132825	0.011803
151	205	02/2017	3597.9	3492.2	3567.5	3645.8	0.008531	0.043989
153	207	06/2017	2719.0	3413.8	3359.8	3174.8	0.190713	0.069998
154	208	02/2017	3421.1	4258.2	3633.9	3738.6	0.058554	0.122023
155	209	02/2017	2861.6	3781.6	3431.4	4188.2	0.166052	0.107512
156	210	03/2017	3573.1	3306.1	3330.0	3298.5	0.073001	0.002297
157	211	03/2017	4203.3	4501.0	4238.1	4723.8	0.008212	0.049505
158	212	03/2017	1793.6	2000.2	1820.1	2031.3	0.014572	0.015554
159	213	03/2017	3323.2	4168.0	3654.6	3558.8	0.090674	0.146156
160	214	03/2017	2967.6	2985.9	3299.4	3147.5	0.100572	0.054114
161	215	02/2017	3300.6	3863.7	3550.7	3826.0	0.070424	0.009749
162	216	03/2017	3352.8	4161.4	3235.2	3836.4	0.036336	0.078103
163	217	04/2017	2946.7	3164.9	2988.0	2884.3	0.013816	0.088656
164	218	03/2017	3890.6	3991.9	4096.9	4004.5	0.0503666	0.0031589
165	219	04/2017	3992	3026.7	3765.9	3443.6	0.0600403	0.1377303
166	220	03/2017	3328.3	4534.1	3595.7	3891.8	0.074356	0.1416533
167	221	05/2017	3182.4	2897.6	3297.7	3076.2	0.0349541	0.0616508
168	222	03/2017	3090.4	3382	3202.6	3450.6	0.0350373	0.020292
169	225	03/2017	3657.4	4122.4	3652	3687.6	0.0014713	0.1054627
170	226	03/2017	2984.3	3874.4	3342.3	3523.2	0.1071038	0.0906513
171	227	03/2017	3226	3325.7	3150	3214.9	0.0241164	0.0333206
172	228	04/2017	3221.9	2669.2	2812.2	2785.7	0.1456988	0.0436562
173	230	04/2017	4411.4	4118.1	4441.4	4048.4	0.0067498	0.0169205
174	231	03/2017	2881.9	3870.9	3517.5	3809.2	0.1806845	0.0159457
175	232	04/2017	3209.9	3480.7	3192.1	3578.5	0.0055763	0.0281085
176	233	07/2017	2670.7	3627.5	3149.7	3412.9	0.1520821	0.0591475
177	234	05/2017	2966.3	3605	3629.3	3421.7	0.1826813	0.0508515
178	235	04/2017	3414.2	3002.8	3370.8	3203.6	0.0128672	0.0668738
179	236	06/2017	3182.6	3975.4	3525.4	3758.5	0.0972321	0.0545625
180	237	04/2017	3586.3	3999.4	3686.9	4004.5	0.0272927	0.0012752

#	PID	Date	model(L)	model(R)	FS(L)	FS(R)	Diff(L)	Diff(R)
181	238	05/2017	2282.9	2401	2445.5	2553.7	0.0664823	0.0636051
182	239	05/2017	3379.5	3000	3422.1	3168.4	0.012459	0.0561309
183	240	05/2017	1575.8	2369.3	1396	2403.1	0.1287736	0.0142855
184	241	06/2017	3570.3	3935.5	3582.5	4096	0.0034086	0.040778
185	242	06/2017	4359.3	4002.9	4492.8	4333.1	0.0297229	0.0825013
186	243	12/2016	2143.8	2548.3	2157.1	2713.9	0.0061663	0.0649853
187	245	05/2017	2999.6	3852.4	3341.1	3903.3	0.1022042	0.0132136
188	250	06/2017	3226.4	4224.2	3482.3	3645.2	0.0734748	0.1370745
189	252	06/2017	2845.7	3859.8	3155.4	3352	0.0981593	0.1315606
190	253	11/2016	3435.5	4340.4	3761.2	4038.4	0.0865906	0.0695763
191	256	06/2017	3385.2	3447.3	2956.1	3570.2	0.1451559	0.0356637
192	257	06/2017	3444.3	3007.1	3088.8	3303.1	0.1151082	0.0984352
193	258	06/2017	3858.1	3699.6	3858.2	3757.3	0.0000191	0.0156066
194	259	06/2017	3966.6	3573	4014.7	3845.4	0.0119782	0.0762515
195	260	06/2017	3233.8	3111.6	2872.3	3175.1	0.125857	0.020416
196	261	06/2017	3004.5	3710.2	3094.4	3193.2	0.0290389	0.139334
197	262	06/2017	3420.6	3003	3534.7	3287.3	0.0322924	0.0946771
198	263	06/2017	3166.2	3333.7	3494.7	3364.6	0.0940104	0.0092632
199	264	11/2016	3686	3315.2	3749	3465.8	0.0168002	0.0454262
200	265	07/2017	3251.7	3358.3	3230.5	3587.8	0.0065732	0.0683462
201	266	11/2017	2746.5	2415.8	2476.9	2534.9	0.108844	0.0492921
202	267	06/2017	3183.1	4048.1	3023.3	3570	0.0528498	0.1180985
203	268	06/2017	2405.7	3284.3	2586.3	2857.8	0.0698109	0.1298532
204	269	06/2017	2738	2981.4	2759.6	3127.1	0.0078094	0.0488859
205	270	07/2017	4865.1	4467.5	4872.6	4933.1	0.001536	0.1042109
206	271	07/2017	3323.7	4412	3746.8	4156.1	0.1129165	0.0580072
207	272	03/2017	3713.8	3536.4	3738.8	3833.1	0.0066928	0.0839004
208	273	07/2017	3480.2	3775.1	3738.8	3833.1	0.0691623	0.0153732
209	274	07/2017	3251	3099.2	3270.5	3324.7	0.005974	0.0727588
210	276	07/2017	3292.6	3634.6	3754.4	4032.7	0.1230027	0.1095302
211	277	07/2017	2201.5	2775.2	2223.9	2665.7	0.0100927	0.0394482
212	278	07/2017	2277	2955.7	2462.5	3040.5	0.0753159	0.0286898
213	279	07/2017	2757.2	2618.3	2851.8	2701.6	0.0331723	0.0318202
214	281	07/2017	3635.1	3553.4	4140.2	3311.3	0.1220068	0.0681403
215	282	08/2017	3719.7	3703.6	3555.9	3508.9	0.0460503	0.0525598
216	283	07/2017	3247.9	3358.7	3496.2	3546.6	0.071031	0.0559502
217	284	07/2017	3045.7	3337.1	3445.3	3503.7	0.1159727	0.0499315
218	285	08/2017	3136.3	2972.1	3118.6	3340.7	0.0056696	0.1240371

#	PID	Date	model(L)	model(R)	FS(L)	FS(R)	Diff(L)	Diff(R)
219	286	08/2017	3803.3	3979.5	3968.2	3964.3	0.0415488	0.0038121
220	287	08/2017	3019.7	2772.8	2635.4	2599.3	0.1458231	0.0625836
221	289	08/2017	3462.9	3171.3	3400.1	3572.5	0.0184566	0.1265199
222	290	10/2017	3012.8	3175	3676.7	3625.2	0.18058	0.141786
223	291	10/2017	2964.8	3828.2	3377.6	3673.6	0.1222247	0.0403885
224	292	08/2017	3088	4437	3838.7	4035.3	0.1955686	0.0905293
225	293	09/2017	3070.4	3522.5	3410.4	3815.9	0.0996913	0.0832806
226	294	11/2017	3101.1	3006	3236.5	3314.9	0.0418315	0.1027666
227	295	10/2017	3339	4163.6	3389.7	4001.7	0.0149712	0.038883
228	298	11/2017	3269.7	3287	3182.3	3358.4	0.0274512	0.0217163
229	299	10/2017	3514.7	3362.5	3627.8	3836.9	0.0311694	0.1410904
230	301	11/2017	3387.1	4210.4	3610.9	3689.4	0.0619851	0.1237484
231	302	10/2017	3167.3	2548.9	2842.8	2728.8	0.1141347	0.0705732
232	303	09/2017	2767.7	2862.5	2842.8	2728.8	0.0264048	0.0466941
233	304	10/2017	3049.3	2599.6	3088.4	2849.3	0.0126733	0.0960368
234	306	11/2017	3482.2	3690.8	3490.5	3550.6	0.0023872	0.0379849
235	307	10/2017	3381	3322.5	3537	3735.5	0.044095	0.1243036
236	309	10/2017	2339.8	2471.1	2346.9	2577.8	0.0030293	0.0431779
237	310	11/2017	3593.8	3639.1	3662.7	3826.1	0.0188093	0.0513982
238	311	10/2017	2871.4	2940.3	2710.6	3079	0.0593243	0.0471844
240	313	01/2018	3607.3	4430.4	4476.4	4305.5	0.1941594	0.0281883
241	314	01/2018	3536.3	4412.8	3501.5	3782.6	0.009926	0.1428116
242	315	11/2017	3736.1	4044.2	3888.9	3866.7	0.0393004	0.0438918
243	316	10/2017	3737	4021.9	3842.3	4593.8	0.0274122	0.1421944
244	317	11/2017	2739.8	3873.5	3215.3	3579.8	0.1478974	0.0758343
245	318	12/2017	3651.3	3862.4	3899.8	4262.4	0.0637192	0.1035712
246	319	12/2017	2808.5	3123.3	2764.1	3154.5	0.0160487	0.0099978
247	320	10/2017	3143.3	4173.7	3832.5	3965.5	0.1798286	0.0498784
248	321	11/2017	3570.5	3541.7	3934.6	3804.4	0.0925337	0.0741631
249	323	11/2017	2393.9	2598.3	2343.6	2543.1	0.0214676	0.0212365
250	324	11/2017	3483.6	4168.6	3362	3575.9	0.0361622	0.1421829
251	325	11/2017	2839.3	3565.9	2811	3190.1	0.0100791	0.1053809
252	326	11/2017	2972.1	3030.3	2809.3	3160.6	0.0579331	0.0429985
253	327	01/2018	4233.8	3911.5	4547.2	3915.8	0.0689301	0.0011042
254	328	01/2018	3411.4	4256.3	3874.8	3691.8	0.1195931	0.1326203
255	329	12/2017	3135.6	3770.3	3233.8	3265.8	0.0303718	0.1338124
256	330	12/2017	2890.8	2682.5	3253.9	3076.1	0.1115889	0.1467322
257	331	02/2018	3885.2	3907.1	4172.2	4331.8	0.0687998	0.1087015
258	333	03/2018	3867.6	3863.2	3675.7	3622.3	0.0522181	0.0623629
259	335	01/2018	3627.3	3771.4	3594.2	3470.4	0.0092104	0.0798195

#	PID	Date	model(L)	model(R)	FS(L)	FS(R)	Diff(L)	Diff(R)
260	338	01/2018	3086.4	3843.2	3594.2	3470.4	0.1412937	0.0969934
261	339	02/2018	3081.7	3293.7	3206.7	3748.9	0.0389709	0.138218
262	340	03/2018	2735.6	3603	3130.4	3170.5	0.1261034	0.1200384
263	341	02/2018	3087.8	3557.8	3015.9	3154.9	0.0238394	0.1132385
264	342	02/2018	3552.8	3301.5	3054.6	3317.1	0.1631114	0.0047168
265	344	02/2018	3393.3	4358.5	3530.8	3713.9	0.0389353	0.1478866
266	345	04/2018	3249.9	3060.2	2979	2863.1	0.0909248	0.0643961
267	346	06/2018	3255.2	3322.9	3023.5	3144.4	0.0766324	0.0537115
268	349	08/2018	3339.9	3185.8	2987.2	3321.9	0.1180751	0.0427358
269	350	05/2018	2995.9	3884.1	2987.2	3321.9	0.0029217	0.1447475
270	351	07/2018	3511.6	4401	3571.7	3986.8	0.0168344	0.0941199
271	352	04/2018	3811.9	3636.9	3930.3	3751	0.0301372	0.0313805
272	353	04/2018	3811.9	3636.9	3930.3	3751	0.0301372	0.0313805
273	354	04/2018	3103.6	4009.6	3557.2	3636.6	0.1275291	0.0930277
274	355	04/2018	3149.8	4063.9	3421.7	3524.4	0.0794605	0.1327617
275	356	06/2018	3234.1	3144.8	3421.7	3524.4	0.0548318	0.1207182
276	357	07/2018	3231.9	3326.7	3301.8	3456.9	0.0211686	0.0391532
277	358	07/2018	2715.8	3422.7	3376.7	3568.8	0.1957144	0.0426948
278	359	11/2018	3870.9	4682.5	4178.9	4565	0.0737082	0.0251009
279	360	06/2018	3256.7	3051.1	2872.3	3175.1	0.1338171	0.040626
280	361	06/2018	2947.5	3189.1	3094.4	3193.2	0.0474841	0.0012966
281	362	07/2018	3302.8	3141.1	3534.7	3287.3	0.0656202	0.0465308
282	363	06/2018	2965.2	3522.8	3488.8	3414	0.1500725	0.030895
283	364	07/2018	3262.6	3419.7	3488.8	3414	0.0648221	0.0016655
284	365	09/2018	3585.2	3948.2	3352.9	3388.8	0.0692907	0.1416854
285	366	07/2018	2652.7	2662.4	2914.9	2428.3	0.0899458	0.0879237
286	368	08/2018	3314.1	4405.9	3662.3	3801.8	0.0950644	0.1371141
287	369	07/2018	3013.5	3116.2	3207.1	3067	0.060362	0.0157987
288	370	07/2018	3894	4858.4	4128.1	4236.1	0.0566969	0.1280851
289	371	08/2018	3442.6	4318.6	3624	3843	0.0500609	0.1101345
290	374	09/2018	3202.8	3904.1	3573.9	3580.7	0.1038282	0.0828294
291	375	01/2019	3214.8	4265.5	3427.7	4085.4	0.0621028	0.0422245
292	376	01/2019	3205.6	4238.8	3518.9	3780.5	0.0890336	0.1081302
293	377	11/2018	3346.7	4062	3744.7	3672.5	0.1062901	0.0958987
294	378	11/2018	2808.4	3727.8	3046.9	3535.4	0.0782747	0.0516046
295	379	09/2018	3505.8	3751.7	4270.1	3792.6	0.1789789	0.0108931
296	380	11/2018	3077.1	3070.3	3276	3259.2	0.0607185	0.0615245
297	381	11/2018	3395.6	4126.2	3652.9	3684.5	0.0704505	0.1070457
298	382	10/2018	3897.2	3769.6	3711.5	3402.7	0.0500244	0.0973289
299	383	10/2018	3088.9	3802.1	3724.7	3419.9	0.1706979	0.1005135
300	384	10/2018	3074.8	3371.6	3724.8	3419.1	0.1745044	0.0141012

#	PID	Date	model(L)	model(R)	FS(L)	FS(R)	Diff(L)	Diff(R)
301	386	10/2018	2919.6	3997.4	3369.5	3530.3	0.1335193	0.116859
302	387	09/2018	3680.4	3370.7	3165.3	3166	0.1627349	0.0607299
303	388	09/2018	2936.4	2788.8	3338.5	2895.3	0.1204311	0.0381863
304	389	09/2018	3355.1	3331.6	3064.8	3507.7	0.0947073	0.0528593
305	391	10/2018	3364.9	3254.3	3064.8	3507.7	0.0979307	0.0778616
306	393	10/2018	2488.2	3709.3	3059.2	3718.7	0.1866356	0.0025466
307	395	01/2019	3102.3	4002.7	3874.7	4164	0.1993365	0.0403051
308	396	01/2019	3283	3014.7	3055.8	3351.3	0.0743372	0.111665
309	397	05/2018	3925.1	3802.8	3725.7	3780	0.0535313	0.0059995
310	400	02/2019	3482.3	4028.6	3059.2	3718.7	0.1382954	0.0769313
311	401	02/2019	3260.9	3789.2	3688.6	3931.1	0.1159521	0.0374593
312	402	03/2019	3361.7	4284.6	3569	4057.5	0.0580764	0.0529936
313	403	05/2019	3001.7	2708.3	2975	3098.1	0.0089723	0.1439266
314	404	04/2019	2901.1	3846.1	3261	3665.9	0.1103574	0.0468555
316	406	04/2019	3431.2	4096.3	3952.3	4483.4	0.1318555	0.0944869
317	409	04/2019	3305.9	3079.4	2870.2	3057.1	0.151795	0.0072504
318	410	04/2019	2825.5	3804.9	3225.8	3356.8	0.1240821	0.1177787
319	411	05/2019	3574.8	3309.7	3945.2	3650.2	0.0938956	0.1028942
320	412	06/2019	3820.7	3798	3716.9	3506.9	0.0279224	0.0766554
321	413	07/2019	3209.3	4330.2	3864.6	4072.1	0.1695589	0.0595971
322	414	05/2019	3611.7	3579.9	3759.9	3715.2	0.039404	0.0377967
323	415	06/2019	2900.2	3956.5	3359.8	3481.5	0.1368043	0.120057
324	418	08/2019	3574.3	4584.6	3827.6	4113.6	0.0661681	0.102733
325	428	01/2020	2477.2	3356.7	2863.6	3118.4	0.1349317	0.0709993
326	430	01/2020	2893.6	3984.1	3129.4	3715.4	0.0753605	0.0674462

VITA

MARYAMOSSADAT AGHILI

2007	B.E., Computer Engineering Amirkabir University Tehran, Iran
2010	M.A., Industrial Engineering and Management Islamic Azad Univeristy NajafAbad, Isfahan
2018	M.S., Computer Science and Information Technology Florida International University Miami, FL
May-Aug 2018	Deep Learning Scientist at Monsanto (Bayer).
May-Aug 2019	Data Scientist at Erricson.
Feb 2020 - Now	Applied Scientist at Microsoft.

PUBLICATIONS AND PRESENTATIONS

Maryamossadat Aghili, Mehdi Shojae, Robin Mayrand, Mercedes Cabrerizo, Naphtali Rishe, Rosie Cid Curiel, David Vaillancourt, Steven DeKosky, David Loewenstein, Ranjan Duara, Malek Adjouadi, "Hippocampus Volume Segmentation Using a Transformer-based Machine Learning Architecture." under review in journal of Alzheimer's and Dementia, 2021.

Aghili, M., S. Tabarestani, M.Adjouadi et al. "Addressing the Missing Data Challenge in Multi-Modal Datasets for the Diagnosis of Alzheimer's Disease ." journal of neuroscience methods, accepted (2021).

Tabarestani, Solale, Maryamossadat Aghili, Mohammad Eslami, Mercedes Cabrerizo, Armando Barreto, Naphtali Rishe, Rosie E. Curiel, David Loewenstein, Ranjan Duara, and Malek Adjouadi. "A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study." NeuroImage 206 (2020): 116317.

Aghili, M., Tabarestani, S., Freytes, C., Shojaie, M., Cabrerizo, M., Barreto, A., ... and Adjouadi, M. (2019). Prediction Modeling of Alzheimer's Disease and Its Prodromal Stages from Multimodal Data with Missing Values. International Journal of Medical and Health Sciences, 13(2), 36-40.

Parvez, I., Aghili, M., Sarwat, A. I., Rahman, S., Alam, F. (2019). Online power quality disturbance detection by support vector machine in smart meter. Journal of Modern Power Systems and Clean Energy, 7(5), 1328-1339.

Maryamossadat Aghili, Parvez, Imtiaz, Arif I. Sarwat, Shahinur Rahman, and Fahmida Alam. "Online power quality disturbance detection by support vector machine in smart meter." Journal of Modern Power Systems and Clean Energy 7, no. 5 (2019): 1328-1339.

Maryamossadat,Aghili, and Ruogu Fang. "Mining Big Neuron Morphological Data." Computational intelligence and neuroscience 2018 (2018).

Parvez, Imtiaz, Maryamossadat Aghili, and Arif Sarwat. "Key Management and Learning based Two Level Data Security for Metering Infrastructure of Smart Grid." arXiv preprint arXiv:1709.08505 (2017).

Azizi, L., Hadi, M., Aghili, M. (2021). Freeway's Traffic Flow Breakdown Prediction Utilizing Disturbance Metrics Based on Trajectory Data. In International Conference on Transportation and Development 2021 (pp. 378-390).

Tabarestani, Solale, Maryamossadat Aghili, Mehdi Shojaie, Christian Freytes, Mercedes Cabrerizo, Armando Barreto, Naphtali Rishe et al. "Longitudinal Prediction Modeling of Alzheimer Disease using Recurrent Neural Networks." In 2019 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), pp. 1-4. IEEE, 2019.

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

Aghili, Maryamossadat, Solale Tabarestani, Malek Adjouadi, and Ehsan Adeli. "Predictive Modeling of Longitudinal Data for Alzheimer's Disease Diagnosis Using RNNs." In International Workshop on PRedictive Intelligence In MEdicine, pp. 112-119. Springer, Cham, 2018.

Aghili, M., Soltani, I. (2011). Designing the Custom Competency Model for Future Business Leaders and Managers (Case Study: Mobarakeh Streel). In Recent Researches in Economics Conference, WSEAS Press, Greece (pp. 116-122).

Maryamossadat,Aghili, and Ruogu Fang. "Towards High-Throughput Abnormal Brain Screening in MRI Images." Women in Machine Learning Workshop (WiML), 2016.

Maryamossadat,Aghili, Malek Adjouadi. "A Refined Deep Learning Model for Optimal Semantic Tiny Object Detection and Segmentation." Women in Machine Learning Workshop (WiML), 2018.