

TIME-TO-EVENT PREDICTION ON ALZHEIMER'S DISEASE USING MULTIMODAL DATA AND MICROVASCULAR BIOMARKERS

BY

DANIEL CHRISTOPHER BIØRRITH, 201909298

MASTER THESIS

IN

COMPUTER ENGINEERING

SUPERVISOR: CHRISTIAN FISCHER PEDERSEN

Co-SUPERVISOR: CHRISTIAN MARIUS LILLELUND

Aarhus University
Department of Electrical and Computer Engineering

June 04, 2024

Preface

This thesis represent the final project of my Master's degree in Computer Science, and is a culmination of 5 years education. I would like to extend my sincere graduation to Christian Fischer Pedersen, my supervisor and Christian Marius Lillelund, my co-supervisor, for their guidance, support and patience throughout the project. I also extend my graduation to the individuals at the Center of Functionally Integrative Neuroscience (CFIN) at Aarhus University, especially the researchers behind the MVAS dataset.

Furthermore, I would like to thank my family and friends for their support and encouragement throughout my studies. Last, but not least, I would like to thank my girlfriend for her patience and moral support throughout the process.

Abstract

This master thesis sets out to perform time-to-event prediction for Alzheimer's disease using survival analysis on two datasets and a combination of both. An extensive state-of-the-art section researches previous work's methods and performances, followed by in-depth literature reviews of Alzheimer's disease, imaging modalities, survival analysis, and data imputation.

The first dataset, Alzheimer's Disease Neuroimaging Initiative (ADNI), is a large multisite longitudinal dataset that contains many data modalities. Much work has been done on the dataset, though few have used the complete dataset for survival analysis. After a thorough exploratory data analysis, the dataset was preprocessed and imputed using Multiple Imputation by Chained Equations (MICE) with Predictive Mean Matching (PMM).

The other dataset, MVAS, was created by the Center of Functionally Integrative Neuroscience (CFIN). It contained unique microvascular biomarkers recently found to be an early factor in Alzheimer's disease. This project sought to explore their effectiveness, but the small size of the dataset made it unable to be used as a standalone dataset. Instead, it would be combined with ADNI3. This dataset was preprocessed and imputed using MICE and PMM.

Four CoxPH models, a classic and three with penalty terms, and a Random survival forest model were trained on five resulting datasets. The models were evaluated using the Concordance Index (C-index), Integrated Brier Score (IBS), and Integrated Partial Weighted C-index (IPWC-C). Results on ADNI Cox models were unable to fit if all subjects were included but performed well in an MCI to AD setting, obtaining a C-index of 0.853. The microvascular biomarkers failed to prove useful, with other biomarkers and cognitive tests outperforming them.

Resumé

Dette speciale havde til formål at udføre overlevelses analyse for Alzheimers patienter på to datasæt, samt en kombination af disse. Først præsenteres en omfattende analyse af de bedst præsterende modeller og metoder, efterfulgt af dybdegående litteraturgennemgange af Alzheimers sygdommen, billedemodaliteter, overlevelsesanalyse og imputering af data.

Det første datasæt, Alzheimer's Disease Neuroimaging Initiative (ADNI), er et stort langtidsundersøgelse datasæt, der indeholder mange forskellige datamodaliteter. Meget forskning er allerede udført på dette, men kun få har brugt det komplette datasæt til overlevelsesanalyse. Efter en grundig undersøgelse af datasættet blev dette forbehandlet og imputeret ved hjælp af Multiple Imputation by Chained Equations (MICE) med Predictive Mean Matching (PMM).

Det andet datasæt, MVAS, er udviklet af Center of Functionally Integrative Neuroscience (CFIN). Dette indeholder unikke mikrovaskulære biomarkører, som for nylig har vist sig at være en tidlige faktor i Alzheimers. Dette projekt forsøgte at udforske deres effektivitet isoleret, men datasættets begrænsede størrelse gjorde, at det ikke kunne bruges som et selvstændigt datasæt. I stedet skulle det kombineres med ADNI3. Det kombinerede datasæt blev forbehandlet og imputeret ved brug af MICE med PMM.

Fire CoxPH-modeller, en standard og tre med penalty term, og en random survival forest blev trænet på fem resulterende datasæt. Modellerne blev evalueret ved hjælp af Concordance Index (C-index), Integrated Brier Score (IBS) og Integrated Partial Weighted C-index (IPWC-C). Resultaterne vist at Cox modellerne ikke kunne tilpasse sig ADNI med alle patienter inkluderet. Derimod præsterede de godt i et MCI til AD domæne, og opnåede et C-index på 0.853. De mikrovaskulære biomarkører viste sig ikke at have nogen forudsigende styrke, og andre biomarkører og kognitive tests præsterede bedre

Contents

1	Introduction	6
1.1	Background	6
1.2	Motivation	6
1.3	Problem Formulation	7
2	Related Works	8
2.1	The use of Machine Learning	8
2.1.1	State of the art	9
2.2	Summary	12
3	Fundamentals of Alzheimer's Disease	15
3.1	Cause of Alzheimer's Disease	15
3.2	A Primer in Brain Anatomy	16
3.3	Biomarkers of AD	17
3.3.1	Macroscopic features	17
3.3.2	Microscopic Features	18
3.3.3	Summary	18
3.4	Diagnosing Alzheimer's Disease	19
3.4.1	Cognitive tests	20
3.4.2	Shortcommings	22
3.5	Vascular Changes	22
3.5.1	Changes in Cerebral Microvascular Blood Flow	23
3.6	Neuroimaging	24
3.6.1	Magnetic Resonance Imaging	25
3.6.2	Positron Emission Tomography	26
4	Fundamentals of Machine Learning	27
4.1	Survival Analysis	27
4.1.1	Models	28
4.1.2	Evaluation	30
4.2	Data Imputing	31
4.2.1	Multiple Imputation	32
5	Implementation and Design	35
5.1	Setup	35
5.2	Alzheimer's Disease Neuroimaging Initiative	35
5.2.1	Exploratory Data Analysis	36
5.2.2	Data preprocessing	42
5.3	Microvascular Biomarkers Dataset	45

5.3.1	Exploratory Data Analysis	46
5.3.2	Missing values	47
5.3.3	Data preprocessing	47
5.4	Combining MVAS and ADNI	48
5.5	Survival Modeling	50
5.5.1	Implementation	50
6	Experiments and Results	52
6.1	Experiments	52
6.2	Results	53
6.2.1	ADNI MERGE	53
6.2.2	MVAS-ADNI	56
7	Discussion, Conclusion & Future Work	64
7.1	Discussion	64
7.2	Conclusion	65
7.3	Future Work	66
Bibliography		67
A	ADNI Dataset	81
B	MVAS Dataset	88

CHAPTER 1

Introduction

1.1 Background

Alzheimer's disease (AD) is the most common type of dementia, accounting for between 60% and 80% of cases. As of 2023, over 50 million people globally have the disease, a number that is only expected to rise and double or potentially even triple by 2050, making it an ever-growing problem [1], [2]. AD is an age-related neurodegenerative disease for which no prevention or satisfactory treatment is currently available. Early and mild disease symptoms might not affect individuals' daily lives. However, the progressive nature of the disease will quickly progress to severe symptoms that hinder the individual's daily life, making them unable to care for themselves. Instead, they require help from caregivers, of which 83% come from family and friends in the United States. The estimated lifetime cost of a person with AD is almost \$400,000 in 2022[3].

Dementia is a general group of diseases with symptoms affecting memory, language, problem-solving, and other cognitive functions. In addition to those, AD has symptoms such as wandering, worsened short-term memory, communication problems, difficulty walking and speaking, confusion, poor judgment, and more. Ultimately, the disease is fatal, with individuals 65 and older typically living 4-8 years after diagnosis.

The disease starts developing up to 20 years before the symptoms start to show. This, coupled with no cure, makes early detection and prevention crucial. However, AD is a highly complex disease with many pathological pathways and is very hard to diagnose effectively, as symptoms can be confused with normal aging. Furthermore, cognitive decline must be present to diagnose AD. The current diagnostic is faulty, and underdiagnosis is a prevalent problem. The increasing number of cases will only further the problem, while there is already a shortage of workforce[3].

1.2 Motivation

Technological advancements throughout the last 150 years have significantly transformed the healthcare industry, introducing technologies such as MRI and PET, and will only continue to do so. One such way is using machine learning and artificial intelligence (AI) to diagnose diseases. Machine learning models have already seen significant advancements in predictive performance. Their ability to generalize and handle high-dimensional data much better than humans shows great potential in the healthcare industry[4].

Much work still needs to be done to incorporate machine learning further. Utilizing it for early detection of AD could help alleviate both the underdiagnosing and increasing shortage of workforce problems and allow for preventive treatment to begin much earlier. Traditional machine learning and deep learning models for binary prediction have been explored extensively, with many studies showing promising results and excellent accuracy[5]. These are supervised learning approaches, however, which limits the use of censored subjects.

On the other hand, survival models utilize all available subjects and predict the time-to-event of AD. Furthermore, they predict the progression of the disease over time. This gives the potential to develop personalized treatment plans based on said disease progression, potentially improving patients' quality of life. Survival analysis has not been explored nearly as much as traditional methods, prompting interest in exploring the area further. With the growing availability of databases containing many different data modalities, research in this area is more relevant than ever. However, the nature of clinical datasets poses challenges to overcome, such as heterogeneous data and missing values.

New evidence suggests that disturbances in the microvascular capillaries of the brain are an early pathological feature of AD[6]. The data from the study, developed by the Center of Functionally Integrative Neuroscience (CFIN) at Aarhus University, has been collected into a dataset known as MVAS that has yet to be employed for survival analysis. The potential of the biomarkers in predicting AD is still yet to be understood. All this leads to the problem formulation.

1.3 Problem Formulation

This thesis aims to design, implement, test, compare, and document a select few time-to-event statistical and machine learning models on 1) the ADNI dataset, 2) the MVAS datasets, and 3) a combination of both. The models shall be evaluated and optimized regarding accuracy, explainability, and uncertainty estimation, with each dataset having a different ordering or importance. The project's goal involves predicting the time to conversion to AD and observing the effect of the microvascular biomarkers on predicting survival prediction. Relevant literature in AD and survival analysis shall be included throughout the process to analyze and further understand the model performances. The models will be evaluated using relevant objective metrics through numerical experiments. Results on the ADNI dataset will be compared against state-of-the-art performance. In contrast, results on MVAS and the combined dataset will be used to assess the potential of MVAS's unique features for survival modeling in the attempt of a proof of concept.

CHAPTER 2

Related Works

On November 4th, 1906, Alois Alzheimer lectured a case study of Auguste D, a 51-year-old woman from Frankfurt whom he had followed for 4.5 years until her death. His lecture is the first time a form of dementia, later known as Alzheimer's disease, was described. Auguste had shown extreme jealousy towards her husband, after which she began developing progressive cognitive impairment in the form of memory impairment, disorientation, hallucinations, and more. Postmortem investigation of her brain revealed a high concentration of protein plaques and neurofibrillary tangles[7]. The decades following the discovery were not very eventful until the 1960s when there was a stark increase in people living till 85 and older, which meant an increase in AD cases. At this point, the diagnosis of Alzheimer's was purely pathological, meaning it was confirmed by examining brain tissue postmortem [8]. This changed in 1984 when it became a clinical, exclusionary approach based on new criteria incorporating biomarkers. These were known as the NINCDS-ADRDA criteria and stood for 27 years. As of this writing, the National Institute on Aging set the newest criteria for AD in 2011, a revision of the NINCDS-ADRDA criteria. They emphasized that the pathophysiological process begins long before the onset of clinical symptoms. Now, AD is diagnosed by testing cognitive ability in combination with biomarkers [9] when available. Some widely used biomarkers include amyloid plaques, phosphorylated tau, neurodegeneration, apolipoprotein-E (APOE) genes, and more. More info on these biomarkers will follow in the subsequent chapters.

There is currently no treatment that cures AD, only treatments that possibly slow down the progression. Approved AD treatments are also sparse; in an 18-year timespan from 2003, no new treatments were approved until Aduhelm in 2021, which was later discontinued [10]. However, as of July 2023, the Alzheimer's Society reported that 141 drugs are being tested in different stages of clinical trials. Among these are immunotherapies, which target the amyloid buildup in the brain, hoping it will slow down the disease pathology. A treatment called Leqembi was approved in the US in 2023, and more are waiting for approval in the US and the EU as of this writing [11], [12]. Multiple companies that research AD also exist. One such is Altoida [13], which develops digital biomarkers by the use of augmented reality and machine learning. They evaluate individuals' cognitive impairments and more to perform precision diagnostics. Another such company is BrainScope [14], who also utilizes AI to help provide diagnostic insights.

2.1 The use of Machine Learning

Intuitively, as there are only treatments to slow the disease, these must be started as early as possible to prolong its progression. For this, early diagnosis is crucial. Despite

the improved diagnosis criteria of AD, the symptoms to definitively diagnose AD are still far into the disease process. The utilization of computer-aided processes may help in this regard. In the last couple of decades, especially since the release of AlexNet in 2012 [15], there has been a rise in the use of machine learning (ML) and deep learning (DL). This is also true for the medical field [4], where artificial intelligence (AI) has brought substantial strides in predicting and identifying diseases. In the field of AD, there have been multiple studies, cohorts, and initiatives that collect data for this purpose, examples of which are ADNI, NACC[16], [17]. These have been widely used, highlighted by the fact that the original ADNI paper has over 1600 citations.

A vast amount of research has been conducted trying to diagnose AD early. Tanveer et al. [5] reviewed 165 papers from 2005 to 2019 that all treated it as a classification problem. Their search split the papers into categories based on the technique. The categories were support vector machines (SVM), artificial neural networks (ANN), and deep learning (DL), with 60, 45, and 60 of each type, respectively. A variety of different methods and modalities have been tested.

Furthermore, the type of problems varied, with most being CN vs AD, and MCI vs AD generally having the lowest accuracy. As future work, they argue that more work in MCI vs AD needs to be further explored. In addition, they observe that more importance has been put on feature extraction than the classification phase. Finally, they argue that more work integrating information from various modalities (multimodal data) is needed, as most use only a single modality (mainly sMRI).

Some of the most widely used modalities include MRI scans of different kinds, PET scans, demographical data, genetic information, neuropsychological and physiological tests (such as MMSE), age, education, and more. Other less traditional non-invasive modalities are also being explored. One such example is by Liu et al. [18], who in 2020 used machine learning on speech data to predict AD.

However, an inherent lack of classification is the absence of information on the progression over time and the inability to handle censored data. Survival analysis is more suitable for this. However, it has not been studied to the same extent as classification and has served as a small subset of the research conducted on AD using machine learning. This is depicted in figure 2.1.

The figure shows that the number of publications on AD using survival analysis is relatively small, albeit it has grown in recent years. These include traditional survival analysis methods, such as the Cox proportional hazard, which often serves as a baseline result, and more modern methods, such as forest-based and deep learning-based methods.

2.1.1 State of the art

The following subsection will present a state-of-the-art review conducted on the most recent research from 2020 onwards. All relevant articles found in the search of figure 2.1b will be reviewed, along with articles found through searching Google Scholar.

In 2020, Spooner et al. [20] provided the most extensive comparison of methods for survival analysis and feature selection to date. They provided a comparison of various machine learning methods for survival analysis, showing that they may obtain high performance compared to the traditional Cox method. They obtained a C-index of 0.93 on

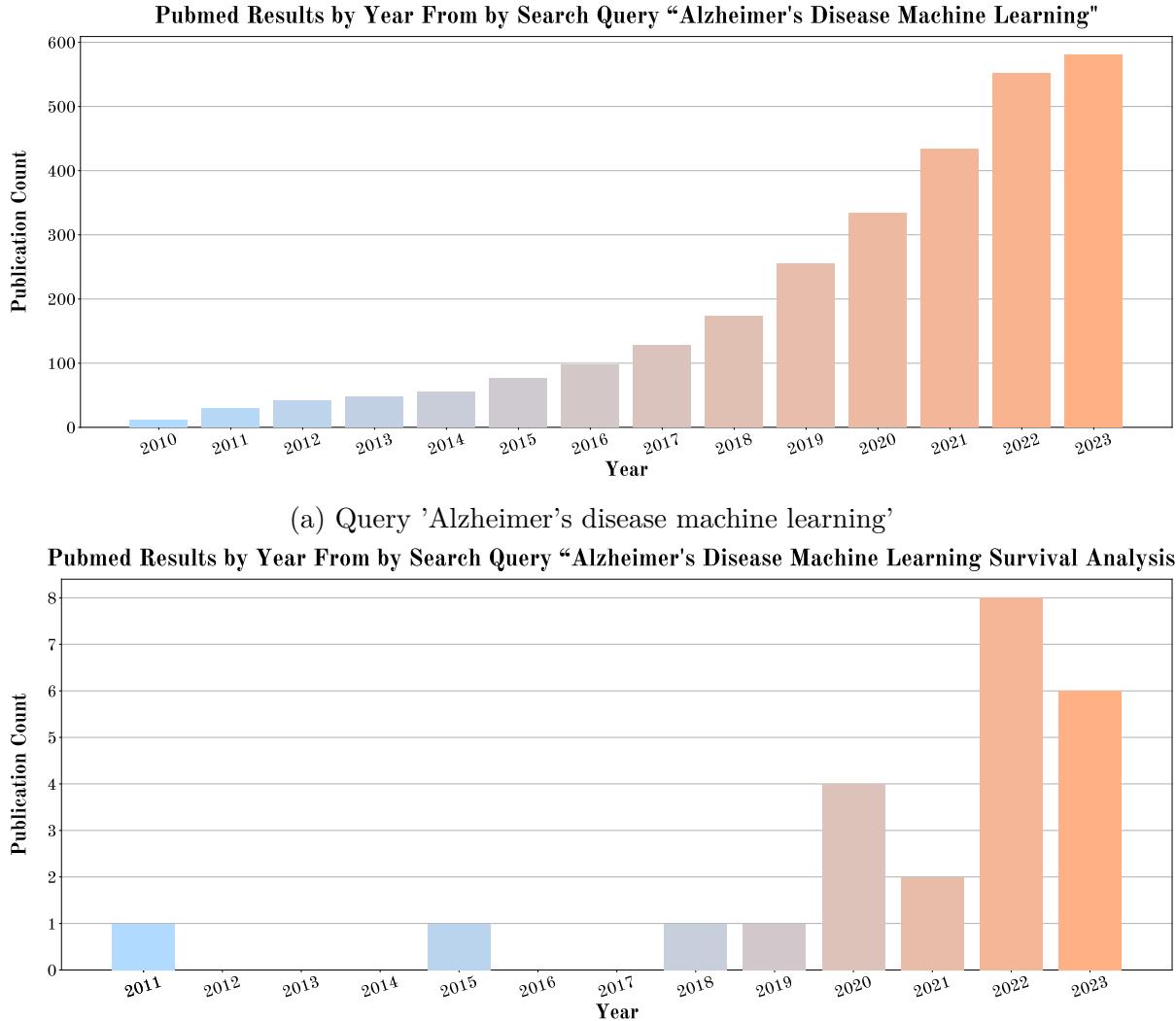


Figure 2.1: Count of the number of hits per year on PubMed searching for only machine learning and searching for survival analysis[19].

ADNI1 using all features in the study. Multiple models, such as CoxBoost, RF-Survival, and ElasticNet, obtained a high accuracy with no statistically significant difference. Additionally, they evaluated feature importance with eight feature selection methods, with which Cox was able to match the C-index of the other methods.

Lee et al. [21] investigated the effects of strokes on the risk of developing AD. They derived brain metabolic signatures representing cognitive impairment with a CNN model trained on FDG PET images from ADNI. The model was transferred to a dataset of patients with strokes. The metabolic signature scores from the FDG PET images were then used to train a Cox model. They found a hazard ratio of 10.12 for the signature score, highlighting the usefulness of FDG PET images in predicting AD post-stroke, but likely also in the general case.

Survival analysis has also been used to compare survival probability between groups. Wang et al. [22] compared the effects of different treatments using the Cox model. They assessed the comparative effect of monotherapy and combination therapies in delaying the

progression of the disease. Biondo et al. [23] compared groups with differing brain-age (an index of brain health obtained from MRI scans) by the use of the Cox model, finding that subjects with a higher brain age had a higher risk of developing AD. Hernández-Lorenzo et al. [24] used biomarkers amyloid- β , fibrillar tau, and neurodegeneration to cluster subjects into different groups. They then used the Cox model to compare the clusters, showing significant differences in the disease progression.

Abuhantash et al. [25] investigates the effect of comorbidities with conversion to MCI as the event of interest. They did this on ADNI1 by utilizing the subjects' medical history. Six different machine learning models and six feature selection models were developed. Of these, a ridge model with permutation feature election achieved a C-index score of 90%.

Aschwanden et al. [26] did a split analysis on the Health and Retirement Study (HRS) dataset to predict the transition to dementia or cognitive impairment from cognitively healthy subjects. First, they performed RSF analysis on half of the data to find the optimal features, which they then used to train a Cox on the other half of the dataset. The significant covariates found included ethnicity (with African American being the strongest predictor), emotional distress, and BMI changes. These were confirmed on the Cox PH models.

Forest based models

Sarica et al. [27] used random forest methods to predict conversion from MCI to AD on ADNI. They used demographical, clinical, and neuroimaging data modalities. The best result was obtained with Random Survival Forest (RSF), obtaining a C-index of 0.87, whereas the Cox model scored 0.83. They later published another paper[28], seeking to improve the explainability of RSF models. They further confirmed that RSF outperforms the Cox model, with a C-index of 0.89 and 0.819, respectively. By using Sapley Additive Explanations (SHAP) to interpret the RSF model, they identified key features such as FDG-PET, ABETA42, and the Hypometabolic Convergence Index (HCI).

Song et al. [29] used RSF on both NACC and ADNI, estimating time to conversion to AD. After performing future selection, they obtained a C-index score of 0.86 on ADNI with only six clinically based features.

Khajehpiri et al. [30] challenged the assumptions of the Cox PH model. They exchanged the linear predictor with the decision tree-based XGBoost model and trained it on the ADNI dataset. Furthermore, they explored the transitions of CN to MCI and MCI to AD. With the nonlinear XGBoost, they obtained a C-index of 0.733 and 0.845 for the CN to MCI and MCI to AD, respectively.

Deep Survival Analysis

Sharma et al. [31] obtained a C-index of 0.79 on the NACC dataset using neural multi-task logistic regression (N-MTLR). Here, they split AD into four stages, and their event was defined as progressing to a later stage.

Nakagawa et al. [32] used deep survival analysis on T1-weighted MRI images to predict time-to-conversion for AD. They extracted grey matter volume for specific brain regions as predictive features, collected from four different datasets. They compared the Cox and DeepHit models, obtaining a C-index of 0.75 and 0.835, respectively.

Wu et al. [33] used DeepSurv on the NACC dataset to predict the stage-specific conversion to AD. Using the CDR cognitive score, they labeled six different stages of

AD, predicting transitions between each stage. They used the Kaplan Meier model and a Log-rank test to compare the progression between the stages. They obtained C-indexes that ranged from 0.86 to 0.91 and an IBS that ranged from 0.06 to 0.94 using DeepSurv on 245 clinical variables.

In 2023, Mirabnahrazam et al [34] performed an in-depth examination of multiple data modalities by deploying a DeepSurv based model on ADNI1. They predicted on subjects in various stages of the disease while comparing the performance of three different data modality sets. Their results showed that no data modality is definitively best for all cases, though cognitive scores of different kinds (e.g., FAQ and CDRSB) had high importance. They analyzed specific features using Permutation Importance to confirm their findings. Using the most important features, they obtained a C-index of 0.831, compared to 0.822 of the Cox proportional hazards method, and an IBS of 0.106 compared to 0.111 of Cox.

Farnsworth von Cederwald et al. [35] employed Bayesian additive regression tree (BART) models on the novel dataset Betula. Betula is a longitudinal dataset that tracks cardiovascular changes. Their results indicated that subjects with an accelerated cardiovascular risk profile had a higher risk of developing AD.

Musto et al. [36] performed a multivariate survival analysis with a wide range of features, including MRI, PET, cognitive assessments, and more from ADNI2. They split the dataset into only CN and MCI patients, with the event of interest being conversion. They implemented a Cox, RSF, and a DeepHit model. The RSF model performed the best, obtaining a C-index of 0.84 for the MCI group and 0.86 for the CN group.

Research using longitudinal and survival data in combination with joint models has also been explored. This allows the use of change in covariates over time instead of only the baseline visit, as is the case in traditional survival analysis. One such model is TransformerJM, a proposed attention-based joint model by Lin and Luo in 2022 [37]. It is based on the transformer architecture and predicts conversion from MCI to AD on the ADNI and NACC datasets. The model takes a sequence of visits as input, from which it predicts the subsequent longitudinal outcome and the survival probabilities, obtaining an IBS of 0.062. However, the data from ADNI was a complete case analysis of only 11 features.

2.2 Summary

A summary of the studies is presented in table 2.1. Notably, many different methods and models obtain high scores in the C-index. Cox model were found to obtain high performance with the feature selection, almost matching the performance of more complex models. However, the RSF method generalizes very well to the high dimensionality and heterogeneity of the data across multiple studies. More emphasis is also being put on deep survival analysis, with more sophisticated models such as DeepSurv and TransformerJM being developed and obtaining good results. The imputation of missing variables has not been a vocal focus point in the studies, with many using simple methods (mean matching, out-of-range value) or complete case analysis. It has generally been a sidenote in most of the studies.

The studies also showed that many different data modalities and types perform well. This includes MRI scans, PET scans (FTP and FDG), demographical variables, and cognitive scores. The importance of cognitive scores in particular was found to be high through feature selection across multiple studies. These include CDRSB, ADAS, RAVLT, and FAQ. Notably, very few have performed survival analysis on the complete ADNI dataset using multimodal data.

Table 2.1: An overview of the implementations and results from related studies on the prediction of AD progression since 2020. Many of the studies have tested multiple models; the best performing are highlighted here

Year	Authors	Datset	Method	Data Modality	Event Target	Imputation Method	IBS	C-Index
2020	Spooner et al. [20]	ADNI1, MAS	CoxBoost	Demographical, genetic, clinical	CN & MCI to AD	MICE	-	0.93
2020	Nakagawa et al. [32]	ADNI, AIBL, JADNI, Shimane	DeepHit (DSA)	sMRI, age and MMSE	CN & MCI to AD	CCA	-	0.83
2021	Sharma et al. [31]	NACC	N-MTLR	Cognitive, demographical	Stage progression	Mean imputation	0.09	0.79
2022	Wu et al. [33]	NACC	DeepSurv	Clinical	Transition between stages	CCA	[0.1-0.06]	[0.86-0.91]
2022	Lin and Luo [37]	ADNI, NACC	Transformer based DSA	Clinical, cognitive assessments	MCI to AD	CCA	0.062 (ADNI)	-
2022	Khejehpuri et al. [30]	ADNI	Xgboost PH	Demographic, MRI, cognitive assessments	CN to MCI, MCI to AD	CCA	-	0.733 (CN to MCI), 0.845 (MCI to AD)
2023	Song et al. [29]	NACC & ADNI;	RSF	Cognitive scores, demographical	cN & MCI to AD	Set to 0	-	0.86 (ADNI)
2023	Mirabnahrazam et al. [34]	ADNI1	DeepSurv (DSA)	Imaging, genetic, CDC	CN & MCI to AD	Replace by out-of-range value	0.106	0.831
2023	Sarica et al.[27]	ADNI	RSF	Demographic, clinical assessments, sMRI	MCI to AD	MissForest	0.11	0.87
2023	Abuhantash et al. [25]	ADNI1	Ridge	Clinical, medical history	CN to MCI	kNN	-	0.90
2023	Sarica et al. [28]	ADNI	RSF	Demographic, clinical, cognitive, CSF, imaging	MCI to AD	MissForest	0.09	0.89
2023	Musto et al. [36]	ADNI2	RSF	Demographic, MRI, PET, cognitive assessments, CSF biomarkers	CN to MCI & MCI to AD	Boolean value indicating missing values	-	0.86 (CN to MCI), 0.84 (MCI to AD)

Abbreviations: **C-index**, Concordance Index; **IBF**, Integrated Brier Score; **CN**, Cognitively Normal; **MCI**, Mild Cognitive Impairment; **AD**, Alzheimer's Disease; **CCA**, Complete case analysis; **CPH**, Cox Proportional Hazard; **RSF**, Random Survival Forest; **kNN**, k-Nearest Neighbors; **DSA**, Deep Survival Analysis; **CDC**, Cognitive tests, Demographic, and CSF; **N-MTLR**, Neural Multi-Task Logistic Regression; **sMRI**, Structural Magnetic Resonance Imaging;

CHAPTER 3

Fundamentals of Alzheimer's Disease

3.1 Cause of Alzheimer's Disease

The cause of most AD cases is still unknown. It is a disease with many pathological features, both microscopical and in later stages macroscopical. It results from multiple accumulating risk factors, as opposed to one single cause [3], [6].

Agewise, AD is divided into two main types, based on when the disease onset. The first is early-onset AD (EOAD), defined as when the disease is diagnosed before age 65, typically in the range of 30-65 years. This accounts for less than 5% of all cases, in which most are familial AD (fAD) [38]. This is characterized by inheriting mutations in APP, PS1, and PS2 genes and typically has a more aggressive course. The mutations all influence a common biochemical pathway: the production of A β (introduced in more detail later). It has a clear molecular background and is easily recognized [2], [7], [39].

The second type is late-onset AD (LOAD), which accounts for more than 95% of AD cases. It is defined as when the disease appears after the age of 65. It has a much more sporadic manifestation and a more cryptic genetic background, though genetic risk factors such as the apolipoprotein E gene (APOE) have been identified.

In addition to the genetic factors, AD is said to be an age-related disease, with the risk of developing the disease increasing with age. Furthermore, it is considered a multifactorial disease, affected by multiple risk factors [38]. This includes the previously mentioned genetic factors, environmental factors, lifestyle choices, and more [2], [40]. A non-exhaustive list of well-known factors is presented below:

- **Genetics and epigenetics**, for example, inheriting one $\epsilon 4$ allele apolipoprotein E (ApoE4) gene triples the risk of developing AD while inheriting two copies of the gene increases the risk by 10-15 times for late-onset disease [41].
- **Lifestyle choices** such as smoking and heavy consumption of alcohol.
- **Pre-existing conditions** such as obesity and insomnia.
- **environmental factors** such as air pollution, diet, and exposure to heavy metals.
- **Medical factors** such as cardiovascular diseases and injuries to the head.
- **Educational length** has also been shown to have an effect on the risk of AD.
- **Gender** has an effect, where women have been found to have a higher risk of developing AD.
- **Ethnicity** also has an effect, where especially African Americans have an increased risk of AD.

3.2 A Primer in Brain Anatomy

Though not extensive, understanding the brain's basic anatomy is necessary for the topic of AD. This section briefly overviews the brain's anatomy, describing significant regions in the disease pathology.

The human brain is a highly complex organ with around 100 billion neurons. It is the center of the central nervous system of humans. Neurons are responsible for processing and transmitting information in the brain via electrical and chemical signals, which means they are quintessential for the brain's working. The brain consists of two main types of tissue: grey matter and white matter [42]. In short, grey matter consists mainly of cell bodies, blood vessels, and neuronal extensions, whereas white matter contains mainly myelinated axons, giving it the white color [43]. Generally, grey matter is towards the brain's outer layers, while white matter lies further towards the center.

While many perceive the brain as one big working mechanism, it is a complex assembly of many regions, each with different purposes and functions. The process of identifying these regions is known as descriptive neuroanatomy [43], [44]. The brain's largest region is the *cerebrum*, which is divided into the left and right hemispheres. The outermost layer of the cerebrum is known as the cerebral cortex and is composed mainly of grey matter. This layer is deeply folded into gyri and sulci, as depicted in figure 3.1b.

Each hemisphere of the cerebrum is split into five lobes, and the four relevant for this project are the frontal, temporal, parietal, and occipital lobes. A brief description and non-exhaustive list of their respective functions is presented below [44], [45].

- **The frontal lobe** is located at the front of the brain. It is responsible for higher cognitive skills such as organizing and short-term memory tasks and contains most of the dopamine-sensitive neurons associated with reward.
- **The temporal lobe** is located at the sides of the brain. It is involved in visual memory, the comprehension of language and emotion, and deriving meaning. Furthermore, it processes auditory information and is also believed to handle short-term memory.
- **The parietal lobe** is located at the top of the brain, behind the frontal lobe. It integrates sensory information from different modalities, particularly spatial sense and navigation.
- **The occipital lobe** is located at the back of the brain. It is the visual processing center that helps recognize colors and shapes.

For a more detailed description of the respective lobes' functions, please refer to box 1.1 in [44]. A depiction of the lobes and their positions is shown in figure 3.1a. An important extension of the temporal lobe is the hippocampus, which plays an essential role in long-term memory and helps regulate emotion and motivation. The entorhinal cortex is located near the hippocampus, inside the temporal lobe. It connects the hippocampus to the neocortex (which will not be introduced here).

Another significant cerebrum region is the diencephalon, in which the thalamus is of interest. It makes up about 80% of the diencephalon and is responsible for filtering, amplifying, and passing desired sensory information to the cerebral cortex. It consists mainly of grey matter [44], [45].

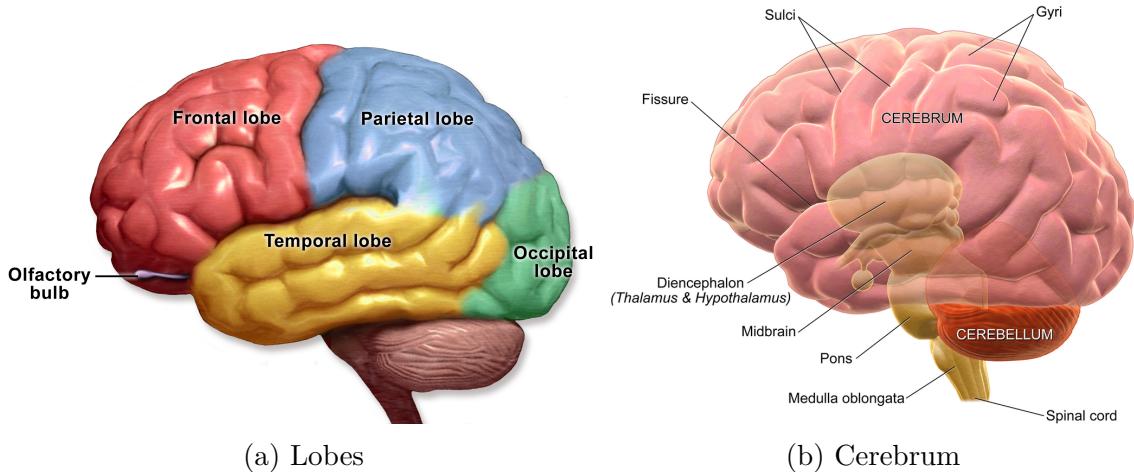


Figure 3.1: Illustrations of the brain and specific regions a subset of regions and loves [46].

In addition to the lobes, each hemisphere contains a lateral *ventricle*. In total, there are four ventricles. The third ventricle is in the diencephalon, while the fourth is inside the midbrain. The ventricles are the communication network with cerebrospinal fluid (CSF) [44].

Finally, two basal ganglia (also known as basal nuclei, many clusters of neurons) regions of interest are found deep in the brain stem. These are the caudate nucleus and the putamen, which are mainly grey matter. They play an important role in habits and motor skill learning and help regulate complex body movements.

3.3 Biomarkers of AD

As AD is a complicated disease with multifactorial pathogenesis, multiple hypotheses have been proposed in an attempt to explain the underlying cause of the pathological changes [47], [48]. Each hypothesis suggests the disease's possible cause(s), where one does not necessarily exclude the other. However, there is no accepted theory for explaining AD's pathogenesis of AD [38]. However, many different micro- and macroscopical features are thought to be associated with the disease. This section provides an explanation of a subset of these features.

3.3.1 Macroscopic features

A visible effect of AD by the use of MRI scans is cerebral atrophy (reduction of the brain). Different regions will simply shrink in volume as an effect of the disease, in a pattern that differs and is much more severe than normal aging. It happens as a result of neurodegeneration. It affects different regions of the brain at different stages of the disease. The medial temporal lobe is one of the regions first affected. The shrinkage of the hippocampus happens early and is one of the most well-established biomarkers. Research has shown that the hippocampus volume in AD patients is 15-40% smaller than in healthy controls. The entorhinal cortex is also affected very early. Progression to mild AD is also found to be associated with a GM tissue loss of 14-19%, specifically a broad area in the temporal cortex. The thalamus area showed a reduction of about 12%, with

80% cell loss of axons. In the basal ganglia, the putamen showed a reduction of about 11% and the caudate between 3-8%. In general, the frontal and temporal cortices are the most affected, with atrophy of the gyri and enlargement of sulcal spaces [39], [49].

Many more regions atrophies, too many to list here, are found to be related to the disease. The result of the neurodegeneration (and, in turn, atrophy) is a reduction in the cognitive power of the brain. None of these are specific to AD but may be highly supportive of it [39], [49].

3.3.2 Microscopic Features

One of AD's believed main pathological features is the abnormal accumulation of Amyloid β ($A\beta$) proteins in the brain. These may aggregate into various assemblies, most importantly plaques and oligomers. Historically, the *amyloid cascade hypothesis* proposed that the aggregation of insoluble $A\beta$ plaques was the main contributor to the pathogenesis of AD. However, later studies demonstrated that they had no direct correlation to the neurodegeneration and cognitive decline that they were first believed to have. Instead, more recent research found that cognitive deficits appeared before the plaque deposition. These findings spawned the *$A\beta$ oligomer hypothesis*, which posits that it is the soluble $A\beta$ oligomers (and not the plaques), that trigger the neurodegeneration that leads to memory impairment and cognitive decline, as well as other symptoms [50], [51].

The disease pathology is also characterized by hyperphosphorylated tau, which is represented by the tau hypothesis. Usually, the tau protein binds and stabilizes microtubules (structures that help guide nutrients and molecules). However, AD causes hyperphosphorylation of tau, which in turn causes them to detach from the microtubules and form neurofibrillary tangles (NFT)[52]. Historically, it was believed that the aggregation of $A\beta$ preceded this, and the hyperphosphorylation occurred because of it. However, more recently, it is believed to be because of a complex interplay between the two, among other factors, though it is still up for debate [53].

Though the accumulation of $A\beta$ and NFTs is found in most AD cases, it is impossible to diagnose AD in the preclinical stages based solely on these features. Healthy individuals have also been found to have plaque and oligomer levels as high as those found in AD patients, without developing any cognitive impairment. The same is true for the hyperphosphorylation of tau[7], [51].

Neuro-inflammation is thought to be another of the main biological processes. It is believed to happen as a result of the brain defending against toxins and injury [2].

3.3.3 Summary

Many different processes occur in the brain in the development of AD. The accumulation of $A\beta$ and Tau are believed to be the main pathological initiators. Other effects include neuroinflammation and neurodegeneration, which, in turn, result in cerebral atrophy. Figure 3.2 shows the current understanding of when the different processes infer.

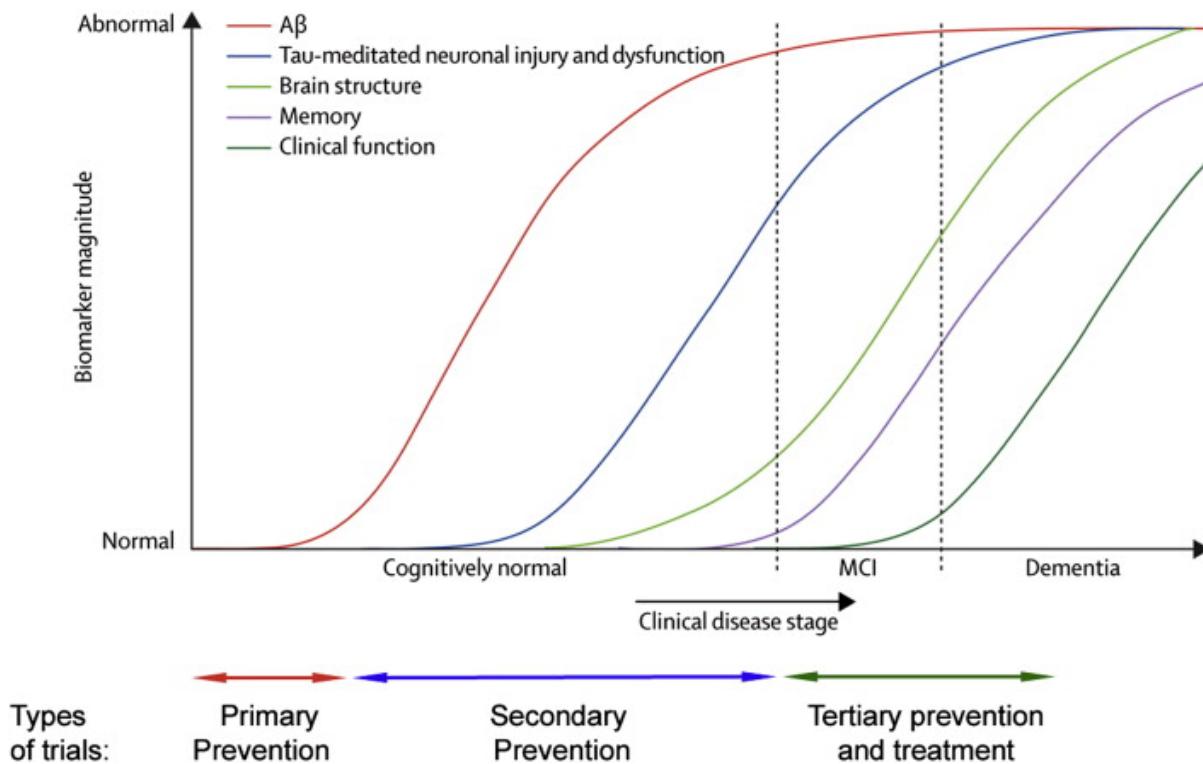


Figure 3.2: The progression graph of Alzheimer's disease[7, p. 323].

3.4 Diagnosing Alzheimer's Disease

As mentioned in section 2, the diagnosis of AD was purely pathological until 1984, when the NINCDS-ADRDA criteria were introduced. These were updated in 2011 by the National Institute on Aging and the Alzheimer's Association (NIA-AA), which still stands as the criteria for diagnosis today. These emphasized that the pathological processes of AD begin many years, even decades, before the onset of the cognitive decline. This introduced a stage between healthy and AD. Instead, there are three main stages of AD progression [7]. These are:

- **Preclinical AD:** Also known as cognitive normal. The pathological processes already begin here, though no observable cognitive change has occurred.
- **Mild cognitive impairment (MCI):** patients with cognitive abilities below their age relatives, but not severe enough to be classified as dementia. Daily functions may be preserved. May also be referred to as prodromal AD.
- **AD dementia:** severe decline in cognitive abilities, to the point where the individual is no longer fully independent. [8].

A depiction of the stages may be seen in figure 3.3. As seen, these MCI may be further divided. This will not be considered going forward.

A cornerstone of Alzheimer's diagnosis is the ATN (amyloid, tau, and neurodegeneration) framework, which supports neuropathological diagnosis. As previously mentioned, these biomarkers have a strong association with AD, though they are insufficient for diagnosis alone, as none are specific to the disease. Furthermore, these are not easily accessible

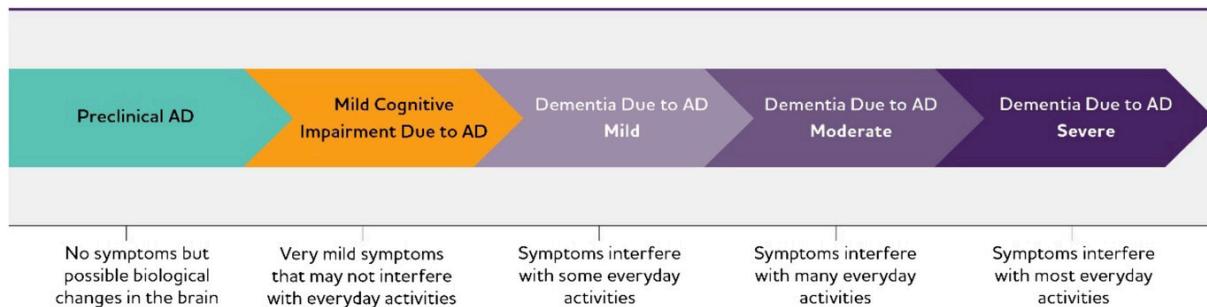


Figure 3.3: Different stages of Alzheimer's disease[8].

to all healthcare providers. The clinical diagnosis of AD is multifaceted and should include a combination of biomarkers, cognitive tests, and clinical evaluation when possible. In addition to the ATN framework, other biomarkers may be used, such as blood tests and inspecting cerebrospinal fluid. Family history, medical history, and assessments of the subject's cognition may also be part of the diagnosis[8], [9], [39].

3.4.1 Cognitive tests

A way to evaluate the cognitive decline of individuals is to perform cognitive tests in the screening processes. These are a way of quantifying a cognitive score based on different criteria. They are an essential tool in the clinical diagnosing of dementia, including AD, and for research. Many such tests exist; a review from 2015 found over 40 different tests available[54], each with different diagnostic performance and focus points. The tests may measure various categories, including memory, language, problem-solving, planning, and more. For the diagnosis of AD, a deficit must be seen in two or more cognitive domains [55]. Thus, cognitive tests are a cornerstone in the diagnosis of AD.

Below is a selected list of important cognitive tests, which includes a brief description of their purpose and possible subtests. The selection is based on the tests present in the datasets used.

- **Mini-Mental State Examination (MMSE):** Though not originally developed for diagnosis, MMSE is a widely used cognitive test developed by Folstein in 1975. It is a short test that evaluates orientation, memory, registration, recall, and more. The result is a score out of 30 (positive), where a threshold is typically set at under 24 for MCI and under 18 for dementia. This may vary, though, based on demographical variables such as age and education, as well as the individual clinic's assessment [55] [56, p. 37-39].
- **Montreal Cognitive Assessment (MoCA):** A 10-minute paper and pencil test that only takes about 10 minutes to complete. It was developed as a screening tool to detect MCI and assess cognitive domains, including attention, memory, language, and more. It provides a score of 30, using a cutoff score below 26 for MCI [57].
- **Clinical Dementia Rating-Sum of Boxes (CDR-SOB):** The CDR is a detailed cognitive test that is used to determine the presence of dementia by evaluating cognitive ability. It quantifies the severity from very mild (score 0.5 or 1), moderate (score 2), and severe (score 3). It rates six domains, which can be summed for the CDR sum of boxes score (CDR-SoB)[58].

- **Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog):** The ADAS was developed in the 1980s to specifically address the cognitive dysfunction in AD. Multiple different versions exist, where we are concerned with ADAS-Cog-11 and ADAS-Cog-13. The number refers to the number of tasks in the test. ADAS-Cog-11 administers 11 tasks and is scored from 0 to 70, where a higher score indicates worse performance. ADAS-Cog-13 is a revised version of ADAS-Cog-11 and includes the two additional tasks of delayed word recall and number cancellation. The score of ADAS-Cog-13 ranges from 0 to 85. ADAS-Cog was developed for studies where cognitive impairments have already happened, which has raised concern regarding its effectiveness in early diagnosis[59].
- **Rey's Auditory Verbal Learning Test (RAVLT):** The RAVLT is designed to evaluate verbal memory. Patients are vocally presented with a list of words that they are to recall over several trials. First, 5 trials of one list are performed, after which one trial on another list is performed, followed by one trial on the first list. From this, multiple scores are obtained. Firstly, the Immediate score is the score for the first 5 trials combined. Thus, a higher score means better cognitive ability. Secondly, the learning score is the score of trial 5 minus the score of trial 1, meaning a higher score indicates better learning and, in turn, cognitive ability. Finally, the forgetting score is the score of the final trial subtracted from trial 5. Here, a high score indicates a poor cognitive ability, as it suggests loss of information over time[59], [60].
- **Functional Assessment Questionnaire (FAQ):** The FAQ is a questionnaire to quantify the performance of daily tasks. It consists of 10 questions regarding topics such as financial management, shopping, and remembering appointments. Each question is scored between 0-3, and the final FAQ score is the sum of those, thus ranging from 0 to 30. A higher score indicates a worse performance[61].
- **Everyday Cognition Scale (Ecog):** The Ecog test is a questionnaire regarding the participant's capability of everyday tasks now compared to 10 years prior. It is split into everyday memory, language, visuospatial abilities, planning, organization, and divided attention. It is based on a 5-point scale, where higher means more cognitive impairment. The participants themselves, as well as a close informant, are asked to fill out the questionnaire[62].
- **Logic memory (LM) delayed recall:** The LM test asks a subject to recall a short story after a delay, typically 30-45 minutes. A higher score indicates better cognitive performance[63].
- **Digit Symbol Substitution Test (DSST):** The digit span score is a widely used paper-and-pen test, wherein subjects are to match symbols to a number. The score is the number of correct symbols placed within a given timeframe (0-125), meaning a higher score indicates better cognitive performance [64]

No cognitive test is perfect, nor may it be able to give a clear, accurate diagnosis. Instead, a combination of tests is preferable to increase the accuracy of the diagnosis. A recent study found that the MMSE and MoCA cognitive tests performed best for the detection of MCI. For dementia, MMSE also performed well, though other tests show similar performance in research [54].

3.4.2 Shortcomings

By now, it is clear that diagnosing AD is a complicated task. It requires many different factors, which naturally increases the risk of human error. Studies show that there is a prevalence of missing diagnosis of MCI and AD, ranging from 25% to 90% [54]. This number emphasizes the importance of even further improvement of methods and finding biomarkers for diagnosing AD.

3.5 Vascular Changes

In addition to the previously stated biomarkers, there is also evidence that suggests the vasculature of the brain is affected by AD. For a better understanding of these changes, a list of key terms is presented below:

- **Cerebral blood flow (CBF):** Also known as cerebral perfusion (passage of blood), CBF measures the rate of blood delivered to a capillary bed in tissue. The standard unit of measurement is milliliters of blood per 100 grams of brain tissue per minute (ml/100g/min) [65].
- **Cerebral blood volume (CBV):** Defined as the volume of blood in a given amount of brain tissue, typically measured in millimeters of blood per 100 gram of brain tissue [66].
- **Mean transit time (MTT):** Is the average duration (in seconds) blood spends within the blood vessels of a particular part of the brain. It is calculated as the ratio of CBV to CBF, meaning $MTT = CBV/CBF$ [67].
- **Blood brain barrier (BBB):** The BBB is a semi-permeable membrane that allows cerebral blood vessels to regulate the movement of ions, molecules, and cells between the blood and the brain, thus helping protect the brain from harmful and unwanted substances that may be found in the blood [68].
- **Tissue oxygen tension (P_tO_2):** A measurement of the partial oxygen pressure in the organ bed. It represents the balance between the local oxygen delivery and consumption [69].

It is well established that BBB dysfunction (increased permeability) and morphological changes in brain capillaries are associated with AD. Furthermore, hypoperfusion in the form of a decrease in CBF, is an early pathological feature of AD [51], [70]. Multiple hypotheses exist in an attempt to explain these changes. Two will be presented: the two-hit vascular hypothesis and the capillary dysfunction hypothesis. Both of them include a vascular contribution to the disease, as opposed to the previously presented hypotheses.

The two-hit vascular hypothesis, proposed by Berislav [71], divides the pathology of AD into two main hits. They are as follows:

- **Hit one:** Vascular risk factors lead to blood-brain barrier (BBB) dysfunction and breakdown and a reduction in CBF. In turn, this initiates a cascade of non-amyloidogenic events preceding dementia. In turn, the BBB provides less clearance of $A\beta$, leading to a buildup of $A\beta$.

- **Hit two:** Increasement and accumulation of A β , leading to accelerated neurodegeneration, dementia, and hyperphosphorylation of tau.

The hypothesis posits a non-amyloid- β pathway (hit 1) and an amyloid- β pathway (hit 2) to AD. The two both independently and synergetically contribute to the neurodegeneration of AD. Most importantly, it posits that the primary vascular factor (and the initiating factor) is hypoperfusion and reduced oxygen delivery.

The capillary dysfunction hypothesis, on the other hand, is based on another model. Said model states that when the capillary transit time heterogeneity (CTH) increases, the maximum achievable oxygen extraction fraction (OEF^{max}) decreases. It is based on an extended flow-diffusion relation (the BKCR equation), which states that regional CBF, capillary permeability, capillary surface area, and CTH limit oxygen availability. According to the model, CBF must be decreased accordingly to maintain oxygen availability when CTH increases. The opposite is also true, meaning an increase in CBF must be accompanied by a reduction in CTH not to reduce oxygen availability [72]. With this, they define **capillary dysfunction** as an increase in CTH during episodes of increased CBF.

Based on the model, the hypothesis suggests that capillary dysfunction, which diminishes oxygen extraction based on the definition above, is the early pathological event in AD. It postulates that the change in CBF (both rising and lowering) reflects an attempt to maintain tissue oxygenation when CTH lowers. While the two-hit vascular hypothesis suggests that hypoperfusion and reduced oxygen delivery are the primary vascular factors, the capillary dysfunction hypothesis suggests that capillary flow disturbances (heterogeneity) and worsened oxygen extraction are the initial vascular factors in AD and even the initial pathological event.

3.5.1 Changes in Cerebral Microvascular Blood Flow

A study by L. S. Madsen et al. [73] further studied the vascular effects of AD. Based on the capillary dysfunction hypothesis, they hypothesized that a deterioration in *microvascular* perfusion is a pathological feature of AD. Specifically, they hypothesized increased heterogeneity in microvascular blood flow and an accompanied prolonging of microvascular transit times.

To test this, Madsen et al. found test subjects from three different groups: healthy controls (**HC**), subjects with suspected non-Alzheimer's pathophysiology (**SNAP**)-MCI and prodromal AD (**pAD**)-MCI. Using dynamic susceptibility contrast (DSC) MRI scans (to be introduced later), they tested the subjects for parameters CBF, MTT, CTH, P_tO_2 , and CBV. These were measured with two DSC-MRI sequences, one sensitive to all vessel sizes (GE DSC-MRI) and one specifically for capillary-sized vessels (SE DSC-MRI) used to measure the microvascular flow. The measurements were performed twice, one for initial data and one at a two-year follow-up.

The GE DSC-MRI results showed small areas of increase in MTT and CTH for the pAD-MCI group and a slight decrease in CBF and CBV, with no noteworthy changes in the other parameters. For the SNAP-MCI group, a slight decrease of CBF and CBV in specific regions was found, with no noteworthy changes in the other parameters. The HC group showed no noteworthy changes in any of the parameters.

The SE DSC-MRI results, however, showed a clear increase in MTT throughout most of the cortex, a decrease in P_tO_2 throughout most of the cortex, an increase in CTH in temporal and frontal areas, a decrease in CBF in frontal and parietal areas, and no noteworthy changes in CBV for the pAD-MCI group. The SNAP-MCI group showed a decrease in CBF, though to a lesser extent than the pAD-MCI group, with no noteworthy changes in the other parameters. The HC group showed no noteworthy changes in any of the parameters.

The finding of changes in MTT, CTH, and P_tO_2 fits well with the capillary dysfunction hypothesis, as no such changes were found in the SNAP-MCI group. This seems to suggest that changes in the microvascular blood flow are a pathological feature of specifically AD. This is in contrast to changes in CBF, which seems to be associated with a more general cognitive decline in normal aging, as it was found in both the SNAP-MCI and pAD-MCI groups (though to a lesser extent), further supported by no observed changes in CBV[73].

Important to point out, however, is that it is not yet determined whether the (micro)vascular changes or the A β and tau aggregations are the initiating factors and primary events of AD. However, the study shows that they all occur in parallel at the early stages of AD, and they appear to be closely connected and may even be reinforcing each other. Microvascular changes seem to be an important pathological feature that is not yet fully utilized.

A possible pitfall of the study to keep in mind is the sparse amount of data, with only 57 subjects having a follow-up, which limits the study's statistical power and possibly its utility for this project. Nevertheless, the data collected in the study created the dataset formally known as MVAS, which has a significant role in this thesis and will be explored in more detail later.

3.6 Neuroimaging

Neuroimaging (imaging of the brain) is a fundamental part of the AT-N framework, as it helps collect key biomarkers and determine the underlying pathological effects and brain evolution. It may help discover cerebral atrophy, (micro)vascular changes, A β and NFTs presence, and more. Many imaging modalities exist, each obtained using different techniques, and each has strengths and limitations citejohnsonBrainImagingAlzheimer2012, islamAlzheimerDiseaseImaging2022.

The two main categories of imaging modalities are MRI scans and PET scans. Each scan may be split into and provide measurements for specific regions. As individuals' brains differ, a common framework for identifying the regions is needed. Neuroanatomical atlases, or simply atlas, are standardized ways to map the brain into regions. Multiple atlases exist, and they all differ in the number of areas, size of the areas, and exact location. This may make it infeasible to compare results from two different atlases without careful consideration[43]. Examples of MRI and PET scans may be seen in figure 3.4¹.

¹Image courtesy of Breton M. Asken and Gil D. Rabinovici, "Identifying degenerative effects of repetitive head trauma with neuroimaging: a clinically-oriented review," Acta Neuropathologica Communications 9, Article number: 96 (2021), licensed under CC BY 4.0.

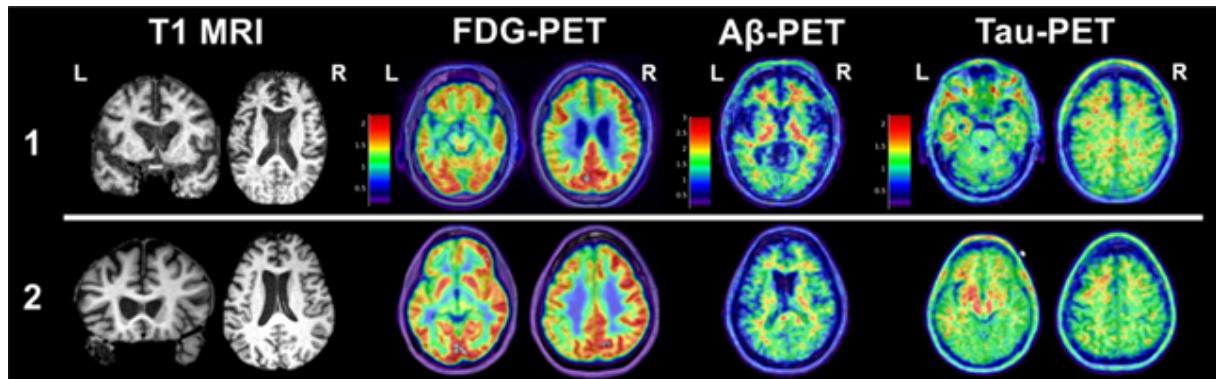


Figure 3.4: Examples of MRI and PET scans. $\text{A}\beta$ was captured with the PiB, and Tau was captured with the FTP tracer. Neither patient 1 nor 2 has AD.

3.6.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a commonly used imaging technique that helps visualize the body's internal organs, the brain's anatomy, and more. MRI is an umbrella term for different, specialized MRI techniques, each providing a different type of information. Generally, all techniques use powerful magnetism (magnetic gradients) and radiofrequency energy pulses/signals to gather information and produce images of the body. Simplified, it exposes hydrogen nuclei (found in abundance in the human body, for example, in water) to a radiofrequency. It then utilizes them by spinning them on their axis and observing the signals created as they return to baseline [74], [75]. The result of this is an MRI sequence that is grey-scaled. The shade of grey represents the signal *intensity*, where white represents high signal intensity, while black represents low [76]. The sequences are high-resolution (0.5-1 mm), and as hydrogen is required, the tissue has to be water-containing [77].

Structural MRI

Structural MRI (sMRI) helps provide information on the brain's shape, volume, and structure. It is widely used for different applications, such as measuring changes in brain morphometry, like cerebral atrophy. Multiple types of specific sMRI scans differ in combinations of pulses and gradients, by which different tissue types have varying intensities on each sequence depending on the combination [78]. The repetition time (TR), the time between application of pulses, and echo time (TE), the time between the pulse and the responding signal after realignment, are varied. Thus, what appears bright depends on how fast a specific tissue realigns [76].

T1-weighted (T1W) and T2-weighted (T2W) MRI sequences are two of the most used pulse sequences employed in sMRI. T1-weighted (T1W) MRI sequences tend to have short TE and TR times. Thus, tissues that quickly realign, such as fat, have a high intensity, whereas liquid realigns slowly and thus has a low intensity. T1W scans are often used to assess changes in volume. Fats and white water/liquids (such cerebrospinal fluid (CSF), infection, and inflammations) generally have a high intensity, muscles have a medium intensity, and grey matter has a low intensity [76], [77].

T2-weighted (T2W) MRI sequences tend to have long TE and TR times, meaning it is the opposite of T1W. Fat has a low intensity, while liquids have a high intensity. Pathologies often involve changes in water content (e.g., inflammation), so TW2 sequences

are used to detect abnormalities and illustrate pathological signatures. Water/liquids (such as cerebrospinal fluid (CSF), infection, and inflammations) and grey matter have a high intensity, while fat and white matter have a low intensity [76], [77].

Dynamic susceptibility contrast MRI

Dynamic susceptibility contrast MRI (DSC-MRI) is a type of MRI perfusion technique, and it is used to give insight into the perfusion (passage of blood) of tissues. It is performed by injecting a bolus (single dose) of a gadolinium-based contrast agent (GBCA). After the injection, rapid T2W imaging (in our case of the brain) is performed, where the effect of the GBCA leads to a signal loss. By utilizing this, values such as MTT, CTH, CBV, and, in turn, P_tO_2 may be extracted from the scan. Once again, different pulse sequences may be employed when performing the scan. In this thesis, gradient-echo (GE) and spin-echo (SE) are the ones of importance. As a key difference between the two, GE DSC MRI is sensitive in all vessel sizes, while SE DSC MRI is relatively more sensitive in capillary-sized vessels, capturing the microvascular biomarkers [79].

3.6.2 Positron Emission Tomography

Positron emission tomography (PET) is a minimally invasive method for measuring metabolic changes using different positron-emitting radiopharmaceuticals (tracers). The tracers usually enter the body by injection, after which it is distributed through circulation. As the positrons are emitted, they collide with electrons, releasing gamma rays in an annihilation event used to generate 3D images[75], [80].

Different types of tracers exist, such as a ligand, which is a molecule that binds to a specific cellular receptor (protein), or metabolic substances (such as glucose or oxygen). The metabolic substance may be used to measure activity in a specific area, as cells will consume the radioactive substance. As activity requires more oxygen or glucose, more tracer is accumulated in the area. For ligand-based tracers, receptor density is assessed by measuring the quantity of tracer ligands bound to the receptors.

Concerning AD, we are mainly interested in finding the buildup of A β and NFTs. PET scans may be performed with specific ligands for these. In the case of A β , one such is ^{11}C -Pittsburgh compound B (^{11}C -PiB, or simply PiB), whereas ^{18}F -Flortaucipir (FTP) may be used for NFTs. In addition, ^{18}F -fluorodeoxyglucose (FDG) PET is a commonly used method for measuring metabolic changes and neurodegeneration in AD.

Important to mention is the fact that PET is minimally invasive (due to radioactive injections) and expensive, with certain limitations, such as the quality of data, which may be poor/noisy. Nevertheless, it has been used as a diagnostic tool for AD and will be included in the modeling.

CHAPTER 4

Fundamentals of Machine Learning

By now, it has been made clear why developing machine learning models to make predictions is a powerful tool in the medical field. Additionally, as presented in section 2, much work has been done in utilizing machine learning to predict and classify AD. However, most of the research is done with traditional machine-learning approaches. These are supervised classification methods that give a singular binary output in the form of a prediction on the patient's diagnosis. Their pitfalls are that they give no information regarding the disease progression over time, and they cannot be used to create a time horizon. If they were to predict the time horizon, they would have to discard subjects that dropped out of the study before the event, as they would otherwise introduce an inherent bias. These cases are formally known as censored observations/subjects and are common in clinical studies. A branch of statistical learning called analysis makes up for these shortcomings, making it of great interest.

4.1 Survival Analysis

The purpose of survival analysis is to predict the risk of an event as a function of time. In other words, it is a model that tries to predict the time until an event occurs, which, in our case, is the diagnosis of AD. Moreover, it provides a method for handling the censored data. Censoring may be split into three categories: right-censoring, left-censoring, and interval-censoring. Right-censoring is the most common type, where the subject exits the study before the event occurs for some undefined reason. Right-censoring will be the sole focus of this project.

A survival problem must be specified for a more concrete understanding of survival analysis. In a study, subjects are monitored for a sequence of observations wherein p features are measured, giving a feature vector $X \in \mathbb{R}^p$. In the realm of survival analysis, these features are known as covariates. A censored subject has a survival time T and a censoring time C . Here, the survival time is the true time of the event occurring, while the censoring time is the time of a given subject's last observation. With these, we may define an event indicator $\delta \in \{0, 1\}$. It is given by:

$$\delta = \begin{cases} 1 & \text{if } T \leq C \\ 0 & \text{if } T > C \end{cases}$$

Thus, if the true survival time T is observed before or at censoring time C , the event indicator is set to 1. If censoring occurs, meaning the true survival time extends beyond the time of the study, the event indicator is set to 0. Finally, we define the observed survival time as $Y = \min(T, C)$. The covariate vector, the binary event indicator, and

the observed survival time may be combined into a triplet (X_i, δ, Y) . This triplet serves as the basis for survival analysis[81], [82].

With the problem definition in place, the goal is to estimate the true survival time T for new observations using the feature vector X . This is typically described by three functions: the survival function $S(t)$, the probability density function $f(t)$, and the hazard function $h(t)$. If one of them is given, the other two can be derived. The widely used survival function is the probability that a subject survives past a given time t , formally defined as:

$$S(t) = \Pr(T > t) \quad (4.1)$$

It is a decreasing function, where a higher value means a higher probability of survival. When $t = 0$, the survival function is 1, whereas when $t = \infty$, the survival function is 0. The cumulative distribution function $F(T)$, the complement to the survival function, is the probability that an event happens before t . Thus, the survival function may also be defined as $S(t) = 1 - F(t)$. The probability density function $f(t)$ is the probability that an individual fails in the very short time interval $t + \Delta t$ or the instantaneous rate of the event at time t . With this, we may define the hazard function, which is the probability density function for T conditional to $T > t$. It is formally defined as:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \quad (4.2)$$

The function is the probability of the event occurring at a very short time interval $t + \Delta t$, given that it has not occurred prior. It may increase, decrease, or remain constant over time. Finally, the cumulative hazard function (CHF) is defined as the integral of the hazard function from time 0 to t :

$$H(t) = \int_0^t h(t) = -\log S(t) \quad (4.3)$$

It is the expected number of events that have occurred at the time t [81], [82]. All these different models are essential in modeling the survival time as a function of the covariates.

4.1.1 Models

Survival models may be split into two main categories: statistical methods and machine learning methods. As a general rule of thumb, the statistical methods are more interpretable, whereas the machine learning models scale better with higher dimensionality of covariates. Hereby follows a non-exhaustive list of popular models.

Cox Proportional Hazard

The Cox proportional hazards model[83] is a statistical regression model used for investigating the effects of the covariates. It is a widely used method and often serves as a baseline for other models. The model is expressed by the hazard function, which it assumes to be mathematically expressed as:

$$h(t, X) = h_0(t) \times \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad (4.4)$$

Here, $h(t, X)$ is the hazard function at time t given the covariates. $h_0(t)$ is the baseline hazard function at time t , which represents how the hazard function changes over time, and it is the hazard function for an individual where all covariates are ignored. It is unspecified, meaning there are no underlying assumptions of its form. X_1, X_2, \dots, X_p are the covariates and $\beta_1, \beta_2, \dots, \beta_p$ are the regression coefficients. The linear combination of X and β is called the log-risk or log-hazard function, while $\exp(\beta X)$ is known as the relative risk or the risk score. The model is optimized by tuning the coefficients to maximize the partial likelihood function.

The model is a semi-parametric method; it makes assumptions about the relationship between the predictor variables and the hazard function. It assumes the effect of the covariates does not vary with time, e.g., different subjects have proportional hazard functions (thereby the name *proportional* hazard). Furthermore, it assumes that each covariate has a linear relationship with the log-hazard function given by the coefficients β . What makes the model semi-parametric is that it does *not* assume a specific parametric form for the baseline hazard function. This is an essential strength of the model as it makes it very flexible and able to model a wide range of relationships between the covariates and the hazard function. However, the model does not necessarily scale well with high-dimensional data and may encounter issues with multicollinearity.[81], [82]

Penalized Cox Models

Introducing a penalty to the Cox model is a common method to improve the model's performance on high-dimensional data. A penalty term is added to the log-likelihood loss function, as a way to prevent overfitting by shrinking the coefficients. The final model retains the proportional hazards form of 4.4. Three commonly used penalizers are Ridge, Lasso, and ElasticNet.

Ridge regression introduces what is known as a l_2 penalty term to the log likelihood function:

$$\log PL(\beta) - \lambda \sum_{j=1}^p \beta_j^2 \quad (4.5)$$

Wherein $\log PL(\beta)$ is the usual Cox partial likelihood function, and λ is a tunable scalar. $\lambda = 0$ results in the standard Cox model, while $\lambda - > \infty$ effectively scales the coefficients to zero. Ridge regression shrinks the coefficients to ensure stability in the model in the presence of multicollinearity.

Lasso regression introduces a l_1 penalty term to the log-likelihood function:

$$\log PL(\beta) - \lambda \sum_{j=1}^p |\beta_j| \quad (4.6)$$

Where $|\beta_j|$ is the absolute value of the coefficients, and λ is a tunable scalar. Lasso differs from ridge regression in that it may also perform variable selection, as it often shrinks coefficients to exactly zero [84]. Finally, ElasticNet is a linear combination of the l_1 and l_2 penalties:

$$\log PL(\beta) - \lambda(\alpha \sum_{j=1}^p |\beta_j| + \frac{1-\alpha}{2} \sum_{j=1}^p \beta_j^2) \quad (4.7)$$

Where λ is the tunable effect of the penalty term, and α indicates the mixing of the l_1 and l_2 penalties. If $\alpha = 0$, the penalty becomes purely l_2 (ridge regression), while $\alpha = 1$ results in a purely l_1 penalty (lasso regression). ElasticNet is known to be particularly useful when the number of features is larger than the number of samples or when there are highly correlated variables[85], [86].

Random Survival Forests

Random survival forests (RSF)[87] is a machine learning survival model that extends upon the random forest (RF) method. The RF method is an ensemble of binary survival trees with introduced randomization in two forms: randomly drawn bootstrap samples and a randomly selected subset of features for splitting. Averaging over the trees with randomization gives robust results for variation in the data and maintains a low generalization error. This has long been a popular method for classification and regression problems, where RSF extends this to right-censored survival analysis. It is a non-parametric method that has been widely used due to its proficiency in handling high-dimensional without having underlying assumptions.

The idea behind the RSF approach is to draw B bootstrap samples from the original data, on each of which a survival tree will be grown. For each tree, p covariates will be randomly drawn. Each node is split using the candidate variables that maximize the survival difference between their daughter nodes. A constraint is applied: each tree may have no less than d_0 deaths, $d_0 > 0$. The CHF is calculated for each tree and averaged to obtain the ensemble CHF prediction. This is finally tested on an out-of-bag test set to obtain a prediction error.

The approach has multiple hyperparameters to tune: the number of bootstrapped trees, minimum number of deaths d_0 , random covariates chosen p , and the minimum number of splits in each tree. There is no clear best choice of parameters; they are chosen through testing.

Random forest, and in turn RSF, is what is known as a black box model. It lacks the explainability we are interested in for assessing feature importance in the clinical application. To combat this, a method called Permutation Importance, suggested by Breiman in the original random forest paper, may be applied [87]. The method involves shuffling values of a single feature and observing its effect on the model's performance. It removes the relationship between the feature and the target and will lower the score if the feature is important.

4.1.2 Evaluation

Obtaining a quantitative score is essential in evaluating a model's performance. Here, we present three different metrics used in survival analysis.

Concordance Index

Harell's concordance index (C-index)[88] is a commonly used metric for evaluating risk score predictions of a given model. It is the probability that the model orders the risk

predictions of two randomly chosen subjects correctly. It may obtain values in the range 0.5 to 1, where 0.5 indicates a model with a predictive ability equal to random, and 1 indicates a perfect prediction. It is obtained by comparing the ordering of the predicted scores to the ordering of observed outcomes for all pairs of individuals that can be ordered. A pair is comparable when one subject has the event occur before the other subject either has the event occur or is censored. A concordance occurs when a subject who experiences the event earlier than another subject is assigned a higher risk score.

For a comparable pair (i, j) , their respective predicted risk scores are denoted $g(X_i)$ and $g(X_j)$. Let $I(\cdot)$ denote the indicator function, which is 1 if the condition is true and 0 otherwise, and let Y_i and Y_j denote the survival time introduced earlier. On the condition that $Y_i = T_i$, Harell's concordance index is mathematically defined as[89]:

$$CI = \frac{\sum_{i \neq j} I(Y_i < Y_j)I(g(X_i) > g(X_j))}{\sum_{i \neq j} I(Y_i < Y_j)} \quad (4.8)$$

However, it has been found that the C-index score will be biased upwards if the amount of censored data is high. A solution proposed by Uno et al. [89] utilizes the inverse probability weighting (IPW) technique, here formally called the IPCW-C score. This approach employs the Kaplan-Meier estimator to determine the censoring distribution. IPW adjusts for censoring by weighting each comparable pair by the inverse probability of being uncensored at the event time. This method reduces the bias introduced by censoring and theoretically provides a more accurate evaluation of the model's performance.

Brier Score

Another metric widely used for evaluating survival models is the Brier score [90]. It measures the mean squared difference between the predicted risk and the observed outcome at a specific time t . The mathematical definition is:

$$BS(t) = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2 \quad (4.9)$$

Where N is the number of subjects, p_i is the probability of event for the i th subject at time t , and o_i is the observed outcome for the i th at time t . The resulting value is in the range 0 to 1, where 0 indicates a perfect model, while a critical cutoff is at 0.25 [91]. Calculating the Brier score at different time points may create a time-dependent curve. Integrating over this gives the integrated Brier score (IBS), which may be used as an overall average performance of the model[92].

4.2 Data Imputing

Data collection in clinical research (such as AD) does not follow a stringent protocol, often resulting in not all variables being recorded for all subjects. This makes the dataset non-heterogeneous, resulting in missing data being a common issue. Machine learning models typically cannot handle empty entries, so missing data must be handled accordingly. What is known as complete-case analysis, which is where subjects with missing data are excluded, introduces bias, reduces variability in the data, and makes it more imprecise[93]. All these characteristics are undesirable. Instead, it is better to predict data

than to discard it, wherein imputation (replacing missing values with plausible values) is needed to preserve the information of the non-missing data.

Before introducing imputation, it is important to understand the types of missing data. A framework developed by Rubin splits missing data into three different categories [93], [94]:

- **Missing completely at random (MCAR):** Missing data is independent of any observed or unobserved variables and unrelated to any characteristics.
- **Missing at random (MAR):** If the probability of the missing data is independent of unobserved variables but instead relates to other observed variables. Thus, they are factually not missing at random but may be estimated from related observed variables.
- **Missing not at random (MNAR):** Also known as informative missing (IM) data, it is when the tendency of the variable to be missing is a function of an unavailable variable. It is when the data is neither MCAR nor MAR.

Additionally, missing data may be univariate or multivariate. Univariate missing data is when a single variable has missing data, while multivariate missing is when multiple variables have missing data. The latter is often the case in clinical research, and handling it is more complex. Many different methods for imputing missing data exist. A traditional imputation technique is mean-value imputation, which replaces missing entries with the mean of the variable[94]. This method, and other methods replacing all missing entries with a single variable, will lead to biased estimates, artificially low variance, and the suppression of possible multivariate relations. Many popular approaches for imputation, such as missForest[95], fail to account for uncertainty in the missing data, essentially treating missing data as if it were real. It will lead to biased results with underestimated standard errors [96]. A method that does address this is multiple imputation.

4.2.1 Multiple Imputation

Multiple imputation (MI) is a widespread method incorporating uncertainty estimation about the true values. The method consists of three main steps, as illustrated in figure 4.1. The first step involves imputing M different values for each missing entry, creating M completed datasets. Here, comparing the variance of values across the datasets may express the confidence level in the imputation. The more a value varies, the less sure the prediction is.

The next step is to analyze the M different datasets using the same analytic methods as if the dataset was complete. Finally, the parameters of the M datasets are combined by pooling to create a single estimation with a given standard error.

MI employs many choices to be made. Austin et al. [94] provided a tutorial covering many of these. How many datasets to impute M is a nontrivial question to which there is no singular answer. It depends on multiple factors, such as how much of the data is missing and the number of variables. Ideally, M should be set so that the standard error of the analysis result does not vary if more datasets are introduced. A rule of thumb is that the number of imputed datasets M should be at least equal to the percentage of missing data. Using between 20 and 100 imputed datasets is common nowadays, as

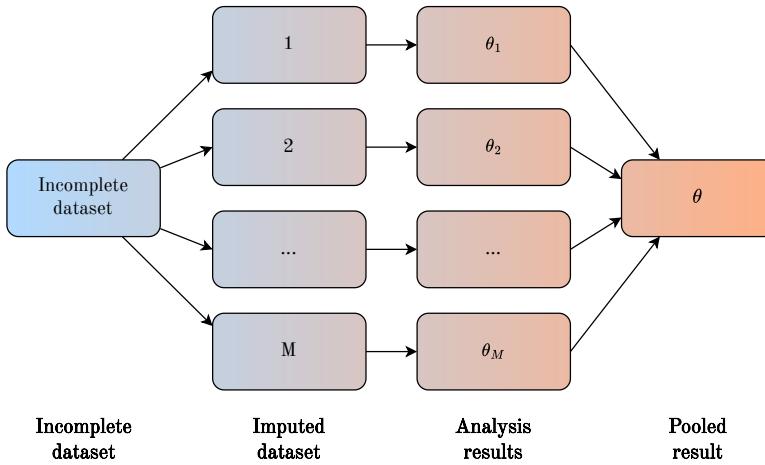


Figure 4.1: Illustration of the multiple imputation process. Inspired by [96].

computation is cheap. Another consideration is choosing which variables to include in the imputation model. Austin et al. recommend including both the time-to-event measure and the event indicator in the imputation, as the estimated regression coefficients for the analysis will be biased towards the null otherwise[94]. Furthermore, they recommend dropping measures with the outcome variable missing, which in our case is the diagnosis.

Multiple Imputation by Chained Equations

MI only sets the framework for imputation, whereas a specific algorithm for the actual imputation must be chosen. Multiple Imputation by Chained Equations (MICE) is a popular algorithm that operates under the assumption that data is MAR. It performs a random draw to fill out missing values and then performs a variable-by-variable imputation. Algorithm 1 summarises the steps[96]. Every variable is regressed on other variables (typically all other variables) to fit an regression imputation model. The imputation model is then used to predict the missing values. It is important to note that the imputation model will be fit both on the observed and on the imputed entries of other variables.

Algorithm 1: A simplified description of the MICE algorithm [94], [97].

Input: Dataset with variables subject to missing data k .

Output: M imputed datasets.

Specify an imputation model for each variable;

for $m = 1$ **to** M **do**

Initialize missing values with random draws from non-missing values;

for $c = 1$ **to** p **do**

for v **in** k **do**

1) Regress the imputation model for variable v on other variables using observed data in v ;

2) Replace missing values using predictions of the model;

Store dataset m ;

A downside to the imputation model is that it uses linear regression, which does not guarantee that the imputed values hold constraints they might otherwise have. This limitation can be problematic for categorical or bounded data. Using predictive mean matching (PMM) combats this. In PMM, a small set of potential donors is made for each missing entry, where the value is imputed by randomly drawing the value of one of the donors. Instead of simply using the best one, making a group of donors helps prevent the overuse of 'good' donors[93]. PMM helps ensure that only plausible values are imputed, which helps keep the restrictions and distribution of the dataset. The number of iterations to run the imputation p is a tunable parameter that should be decided from observing convergence.

Studies have found that MICE performs among the best on clinical datasets. Furthermore, it allows for flexibility in the imputation model, as it can handle different types of variables [94], [98], [99].

Model Pooling

Upon training the models, each of the obtained statistics of interest $\theta_i, i = (1, \dots, M)$ estimated on the M datasets must be pooled into one model. It includes estimating the statistics themselves, as well as their variance. Rubin created a set of rules for handling this. A statistic of interest θ is pooled by averaging the statistic across the M datasets, defined as:

$$\theta = \frac{1}{M} \sum_{i=1}^M \theta_i \quad (4.10)$$

The variance of the statistic, $\text{var}(\theta)$, is calculated using the estimated within imputation variance W_i of θ_i and the between imputation variance B of the estimated statistic. The within-imputation variance is the mean of the variance of the M estimates, given by:

$$W = \frac{1}{M} \sum_{i=1}^M W_i \quad (4.11)$$

The between-imputation variance of the estimated statistic is given by:

$$B = \frac{1}{M-1} \sum_{i=1}^M (\theta_i - \theta)^2 \quad (4.12)$$

Using these, the variance of θ may be calculated by:

$$\text{var}(\theta) = W + (1 + \frac{1}{M})B \quad (4.13)$$

The variance is a quantitative measure of how the statistic varies across the M datasets [94], [96]. Relating this to survival modeling, pooling of Cox regression has already been documented [100]. However, no straightforward way of pooling RSF models exists due to the form of decision trees, and no literature covering the topic could be found. Intuitively, combining the forests would simply result in a new larger forest instead of a pooled model without any uncertain information.

Implementation and Design

5.1 Setup

Figure 5.1 shows the complete process pipeline. It consists of four steps: data pre-processing, survival modeling, pooling the models, and testing. The implementation is done in Python using the libraries Numpy[101], Pandas[102], and scikit-learn[103] for data preprocessing, while the modeling was done with scikit-survival[104] and Lifelines[105]. This section covers the first two steps: data preprocessing and survival modeling. Data preprocessing is split up into preprocessing of ADNI and preprocessing of MVAS.

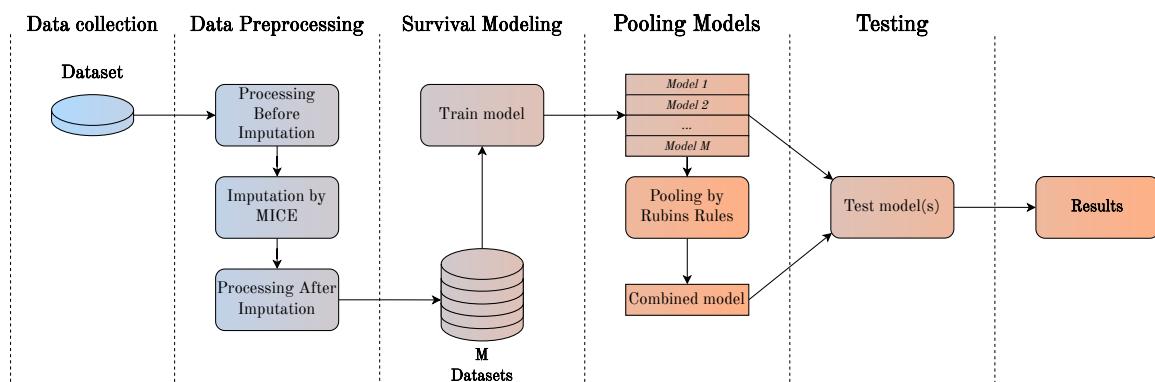


Figure 5.1: Pipeline of the whole process.

5.2 Alzheimer's Disease Neuroimaging Initiative

Alzheimer's Disease Neuroimaging Initiative (ADNI) is a global research study launched in 2003. The goal of the study is to track the progression of AD by using clinical, genetic, imaging, and biospecimen biomarkers and ultimately validate the biomarkers. The study included patients at all stages of AD and healthy controls. It is a multisite longitudinal study that collected data across four different cohorts: ADNI1 (2004-2010), ADNI GO (2009-2011), ADNI2 (2011-2016), and ADNI3 (2017-2022)[106]. As of this writing, ADNI4 is underway, but as it is not complete, it will not be used in this study. ADNI is one of the most widely used open-sourced databases, as shown in chapter 2.

The nature of data collection in clinical research and the multivariate dataset spread across multiple cohorts creates specific problems. It means that the dataset is heterogeneous and has many missing values, which requires imputation. However, before the imputation, an exploratory data analysis must be performed. This is done to understand the data's underlying structure, characteristics, and distribution.

5.2.1 Exploratory Data Analysis

ADNI provides a collection of many different features spanning many different tables. To help combat this, the creators have made a table that collects key features called ADNIMERGE. It comprises the most important biomarkers, demographical variables, cognitive tests, diagnoses, and timestamps. Please refer to table A.1 in the Appendix for a complete list and description of all the features. ADNIMERGE will be the primary source of data for this project. The table comprises 116 different features, 16421 observations, and is spread across 2430 subjects. The features may be divided into three high-level groups:

- **Demographical variables** such as age, gender, and education.
- **Cognitive test scores** such as MMSE, ADAS11, ADAS13, RAVLT, FAQ, and MOCA.
- **Biomarkers** such as APOE4, ABETA, and volume measurements of specific brain regions.

Besides these, the dataset contains features with general information such as the cohort, subject ID, site, diagnosis, and timestamp. Finally, 49 of the features are the baseline measures. These are discarded for future analysis, as their representation through the other features makes them redundant.

Cohorts

The number of subjects varies for each cohort. The distribution of observations across the cohorts, regarding the number of subjects and measurements, is shown in table 5.1.

Dataset	Number of subjects	Total measurements
ADNI1	819	5013
ADNI2	790	4718
ADNIGO	131	382
ADNI3	690	2044

Table 5.1: Number of subjects and measurement of each cohort in ADNI merge. Overlaps between the cohorts are not included.

The table shows that the largest cohort in both subjects and observations is ADNI1, followed by ADNI2, ADNI3, and finally ADNIGO. Some subjects were also enrolled for subsequent cohorts, meaning there is some overlap between them. The number of subjects and measurements in these overlaps are shown in table 5.2.

The table shows a considerable number of subjects that overlap between cohorts. It is important to note that the table shows subjects in over two cohorts multiple times. 727 unique subjects out of 2430 (roughly 30 percent) are present in at least two cohorts, accounting for 4264 measurements.

Missing values

Examining the spread across the cohorts is especially important because each cohort was created with a different goal and at different points in time. This means there is a

Dataset	Number of subjects	Total measurements
ADNI1 to ADNI2	227	1516
ADNI1 to ADNIGO	210	422
ADNIGO to ADNI2	119	711
ADNIGO to ADNI3	50	165
ADNI1 to ADNI3	90	316
ADNI2 to ADNI3	305	1134

Table 5.2: Overlaps between the cohorts in ADNI merge. Users may be counted multiple times if they are present in multiple cohorts.

discrepancy between them in the features recorded, meaning compatibility is not a given. Table 5.3 shows the most subset of features not included in all cohorts and which cohorts they are missing in. A complete list of all affected features may be found in table A.2 in the Appendix.

Variable	Missing in cohort
PIB	ADNI2, ADNIGO, ADNI3
AV45	ADNI1
FBB	ADNI1, ADNIGO, ADNI2
ABETA	ADNI3
TAU	ADNI3
PTAU	ADNI3
DIGITSCOR	ADNI2, ADNIGO, ADNI3
MOCA	ADNI1
All Ecog measures	ADNI1

Table 5.3: List of variables missing in cohorts.

The difference in cohorts also results in the distribution of missingness varying across them. Figure 5.2 highlights this by showing the distribution of missing percentages per feature for each cohort.

The figure illustrates that the missingness inside each cohort has a cutoff around 50-60 %, with most of the features being under. However, small spikes at high percentages (>80%) are present, especially in ADNI1. This is mainly due to the number of features not included in the cohort, as seen in table A.2. Figure 5.3 visually represents the missingness across all cohorts. The figure shows the values not observed in specific cohorts, such as Ecog measures in ADNI1. Generally, the missingness appears random and without patterns, though with exceptions. One such is Ethorinal, Fusiform, and Midtemp measurements in ADNI1, which are missing in a large patch.

Previous research has found that ADNI1 is MAR [107]. This is also an assumption held by other related studies, such as the one by Spooner et al. [20]. The strict data protocol for the data collection helps ensure this, even though there are many missing values. It is assumed that other observed features may explain the missingness. Examples of this are the time of measurement, which may help explain missingness in features only observed at baseline, or the cohort as. However, it is not a given that this applies to the other cohorts. In figure 5.3 it's clear to see that the collection in ADNI1 is more strict, especially towards the last half, compared to the other cohorts. However, for this project, it's assumed that the data is MAR across all cohorts, using the same arguments as for ADNI1.

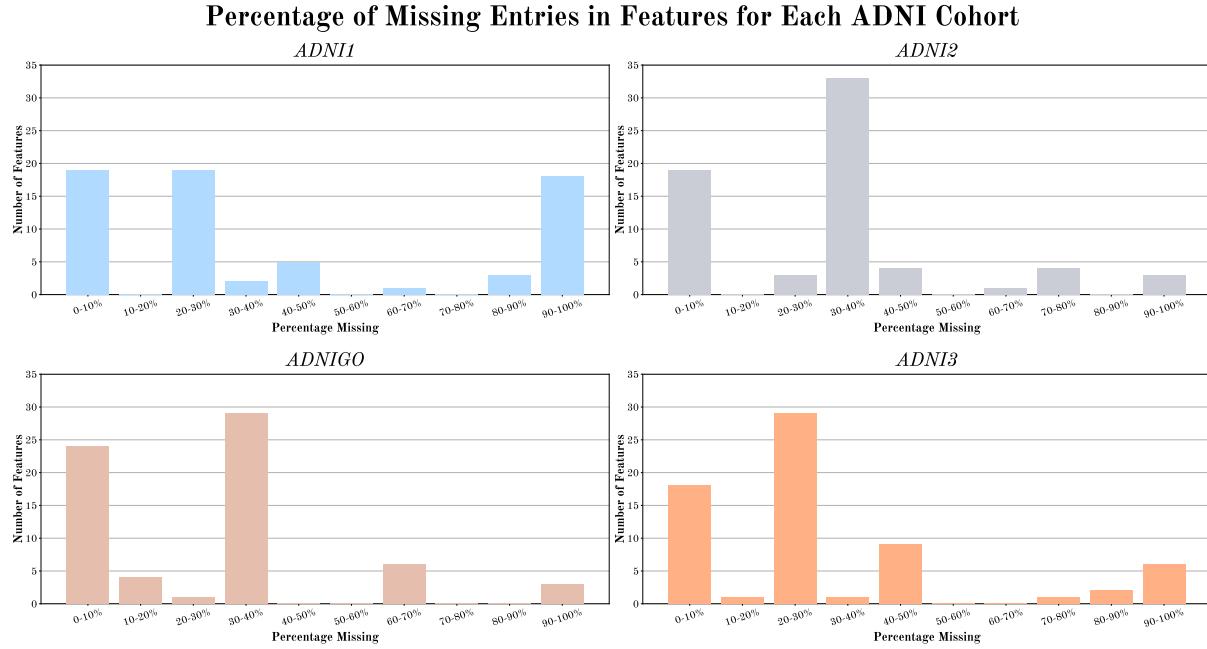


Figure 5.2: Overview of missing values in individual cohorts of ADNIMERGE.

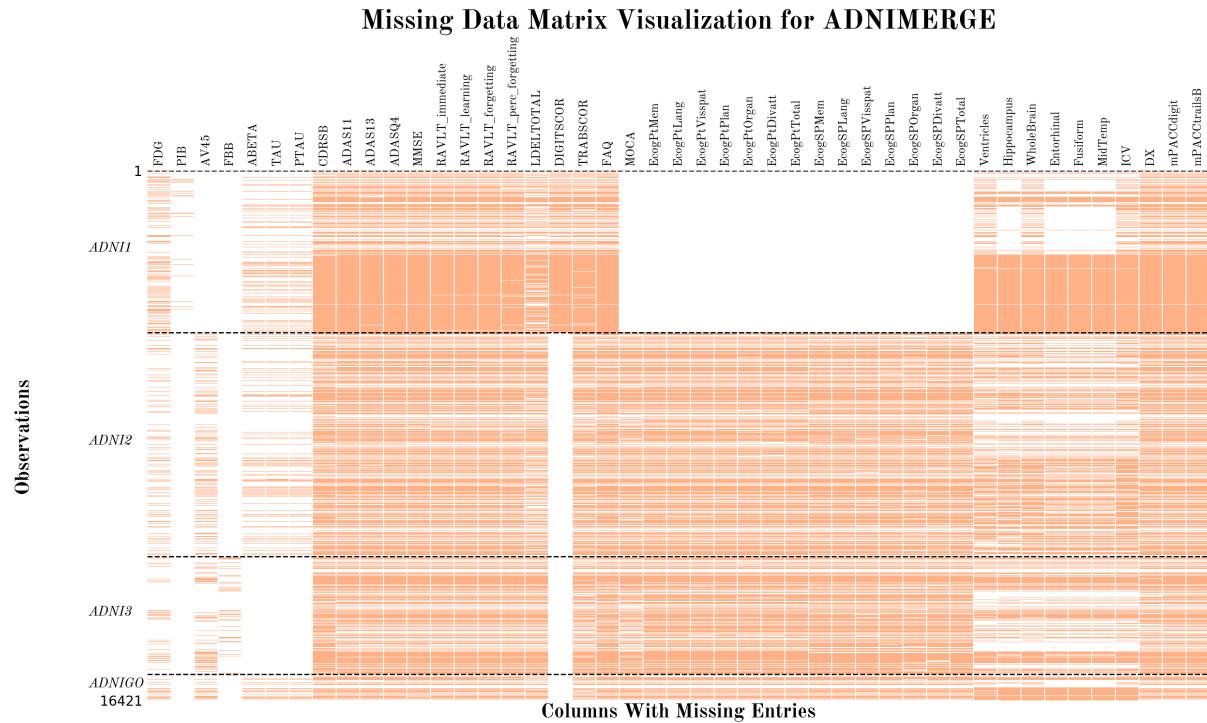


Figure 5.3: Missing data matrix of ADNIMERGE.

Diagnosis

The 'DX' feature is the diagnosis given to the subject at each visit. It is a categorical variable: AD, MCI, or Dementia, where dementia means having AD. For the sake of clarification, Dementia is changed to AD. The distribution of the baseline diagnosis is shown in figure 5.4. The majority of subjects are either CN or MCI at baseline. However, 413 subjects start with a diagnosis of AD, making them ineligible for survival analysis.

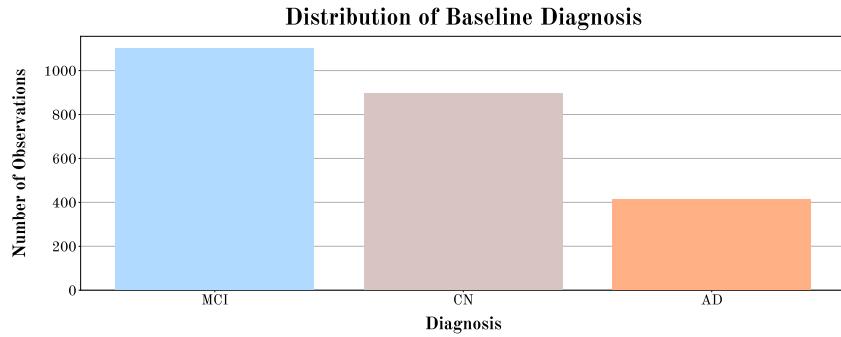


Figure 5.4: Distribution of baseline diagnosis in the ADNIMERGE dataset. 11 subjects don't have a diagnosis at baseline.

Due to the nature of survival analysis, the change in diagnosis over time is also of great interest. A Sankey diagram was created to help visualize this, as seen in figure 5.5. It shows the change from the diagnosis at the baseline visit to the diagnosis at the last visit of each respective subject.

Sankey Diagram of Subject Diagnosis (DX) Transitions

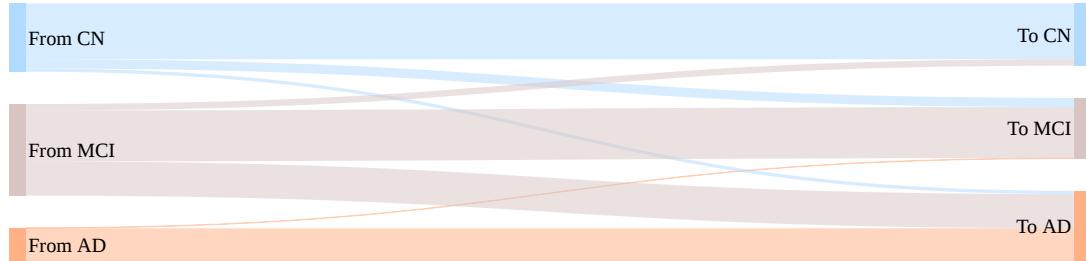


Figure 5.5: Sankey diagram of transition between different diagnoses.

The exact number for each transition may be found in table 5.4. Changes in the diagnosis across the dataset do occur, though the number of changes to AD is relatively low. Specifically, 404 subjects converted to AD, accounting for around 16.5% of the total subjects (excluding subjects starting with AD).

	From CN	From MCI	From AD
To CN	601	69	0
To MCI	99	550	5
To AD	36	368	358

Table 5.4: Transitions from baseline diagnosis to the last diagnosis for each subject.

From the Sankey diagram, it's also seen that cases exist where subjects with a diagnosis of AD regress to MCI. This prompts a further inspection, as it may be a sign of misdiagnosis being prominent in the dataset, which could affect the model negatively. If the AD diagnosis is a case of false positive, the event for the survival analysis hasn't occurred. In the inspection, where all empty diagnoses were removed, 26 subjects were found to at one point have been diagnosed with AD and later regressed to either MCI. A

manual inspection of how each of the subject's diagnosis changes over time was performed through visual plots. A subset of the plots may be seen in figure 5.6.

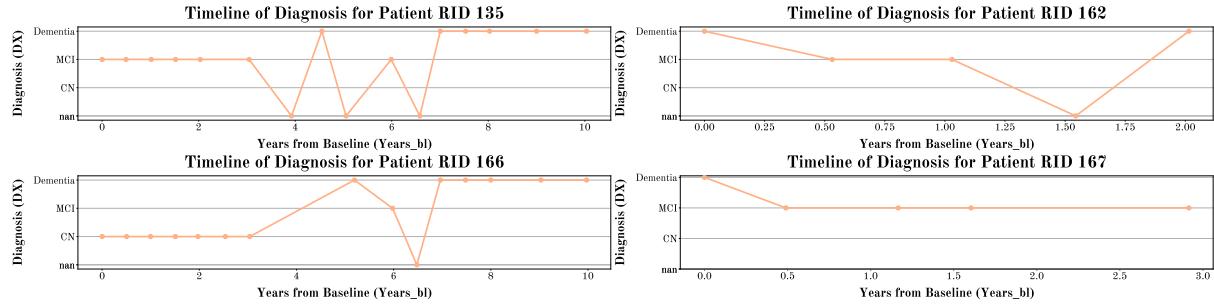


Figure 5.6: A subset of the plots of diagnosis over time for subjects who get diagnosed with AD and then back to another diagnosis.

For plots of all cases, please refer to figure A.1 and A.2 in the Appendix. From the plots, it is clear that the diagnoses are inconsistent and could negatively affect the model if not handled in the preprocessing.

Unfortunately, the 'DX' column also has 4963 empty entries. Roughly 30% of the diagnosis entries are missing, a rather significant amount. This might make the event time less precise, which is unwanted. Inspecting the distribution of empty diagnoses, shown in figure 5.7, shows clear spikes at time points in between whole years, such as 3, 18, and 30.

This is likely explained by the clinical study schedule of ADNI, where certain features are only measured at distinct visits [108]. The results of this will also introduce bias towards events and censoring occurring at time points representing a whole year, such as 12, 24, and 36. This is confirmed in figure 5.8, which shows the distribution of time points for events and censoring. It shows clear spikes at time points for whole-year follow-ups. It also shows that the largest censoring times far exceed the largest event times, which will have an impact on the Cox models, setting the baseline hazard to zero at those time points.

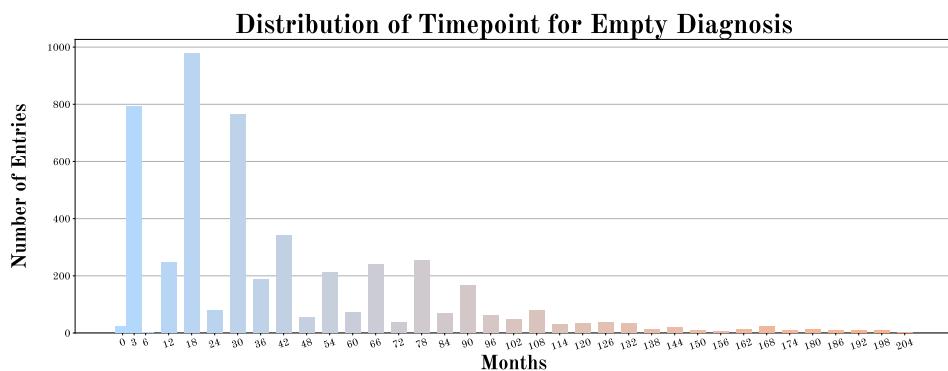


Figure 5.7: Distribution of timepoints of missing diagnosis.

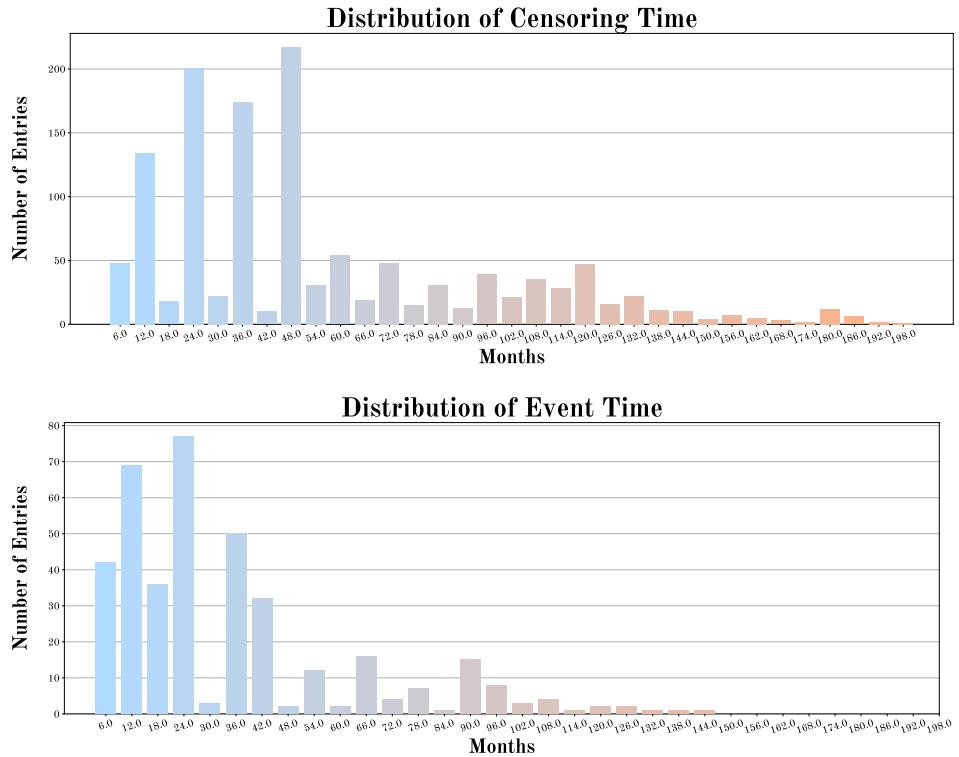


Figure 5.8: Distribution of censoring and event timepoints.

Distribution of visit timepoints

Another important aspect to inspect is the distribution of the visit times across time points. Figure 5.9 illustrates exactly that. Around 90% of visits happened at or before month 90, while around 80% happened at or before month 60.

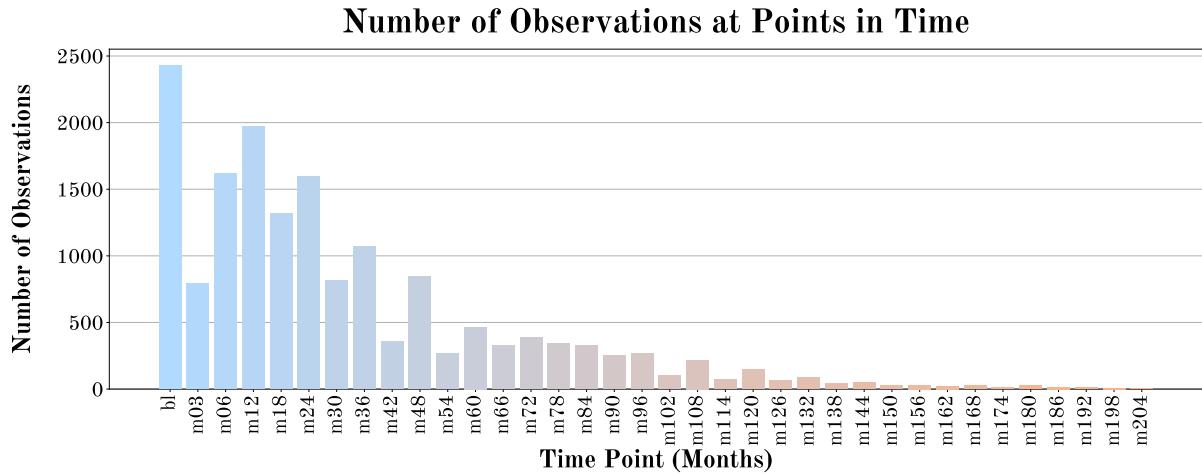


Figure 5.9: Distribution of measurements at time points in ADNIMERGE.

Figure 5.10 shows the time distribution for each subject's last visit. Around 50 % of subjects had their last visit at or before month 36, 75 % at or before month 60, and around 90 % at or before month 120. Additionally, it shows clear spikes at months 12, 24, 36, and 48.

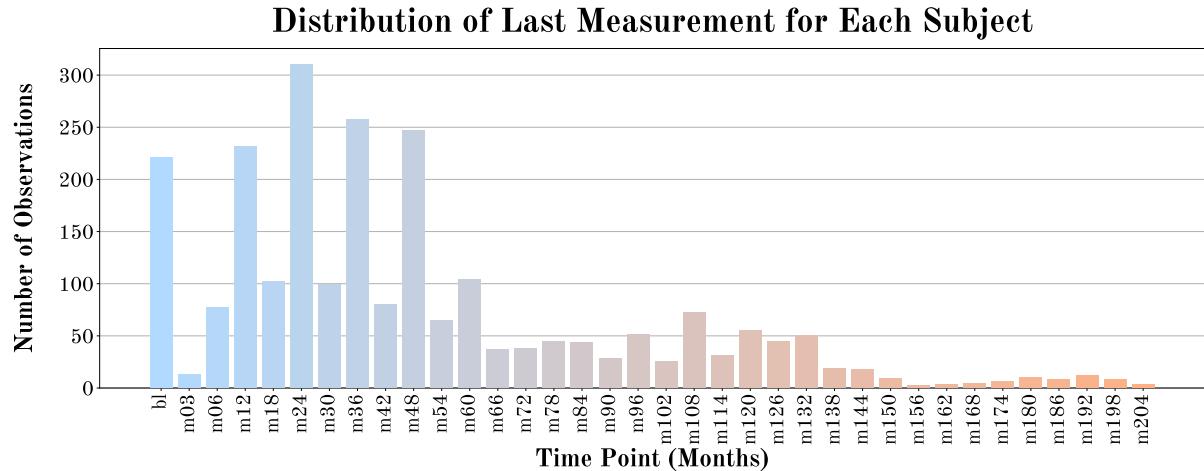


Figure 5.10: Distribution of the last measurement for each subject in ADNIMERGE.

These two figures show that the majority of the visits occur within the first year and a half years of the study and a vast majority within the first five years. Considering the long pathological progression of AD, this is relatively low.

Categorical variables

Finally, the categorical variables to be used in the analysis must be inspected. Figure 5.11 shows their distribution.

Figure 5.11a, shows the spread of APOE4 genes, where the majority have zero or one gene, and few have 2. Figure 5.11b shows that a vast majority have over 12 years of education. Figure 5.11c shows that the majority of the subjects are not Hispanic/Latino, while 5.11d shows the majority are white, with a small spread across the other races. Figure 5.11e shows a near-even gender split, though with more males. Finally, figure 5.11f shows the spread of marital status, with the majority being married. With the distribution in race and marital status, it may be beneficial to combine some categories.

5.2.2 Data preprocessing

Data preprocessing is necessary for the dataset to be applicable for survival analysis. This includes handling missing values, disregarding irrelevant and/or redundant features, how to handle categorical variables, determining events, and more. The preprocessing is split into two main steps: preprocessing the data and imputing data.

Before imputation

For the dataset to be appropriate for imputation, it must be formatted correctly. This is done to make imputation possible and help obtain the best possible results. The formatting builds upon the findings of the exploratory data analysis, as well as the theory from chapter 3.

The first important aspect is how the missing diagnosis entries are to be handled. The choice was made to remove the missing entries following the guidelines of [94] regarding missing entries in the event of interest. Leaving the diagnosis to the imputation model would be too imprecise, as it would depend on simple linear regression. Finally, a test

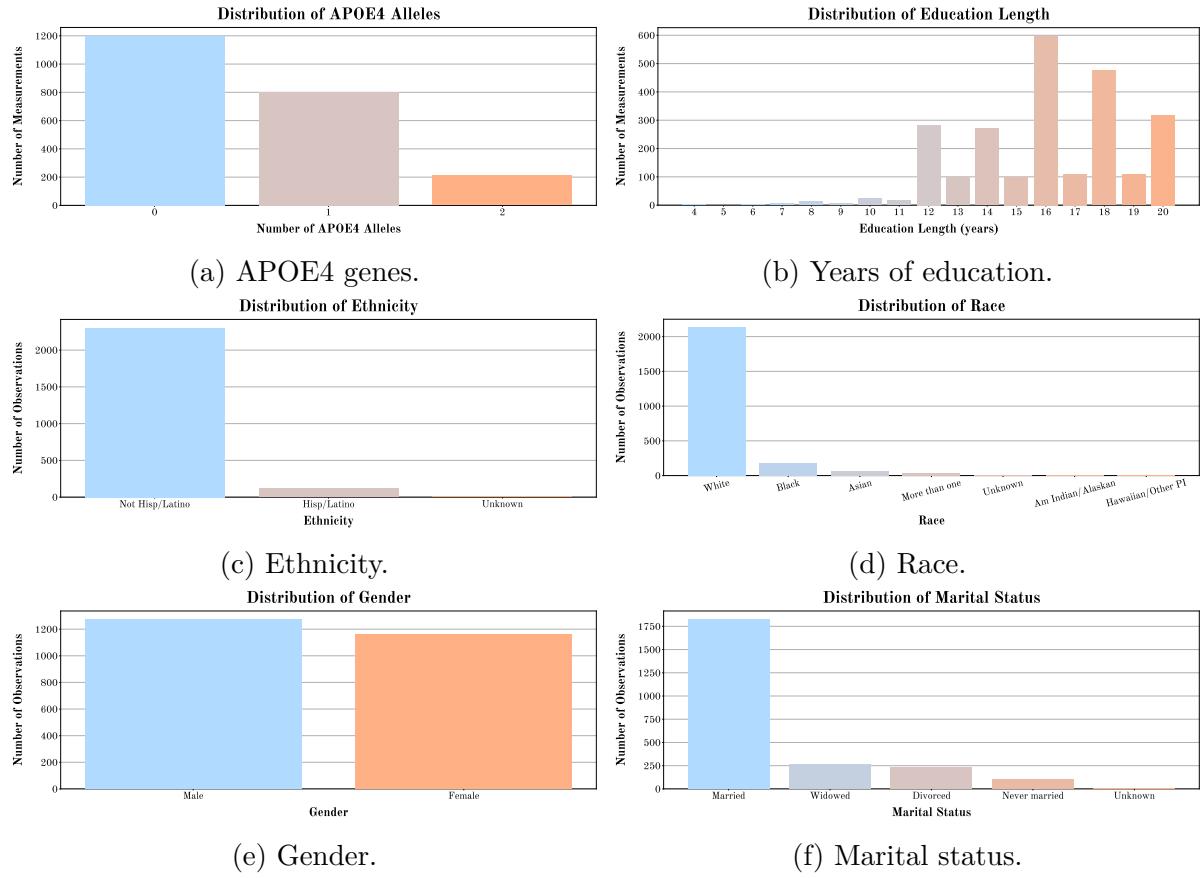


Figure 5.11: Distribution of the categorical variables in ADNI.

was conducted by inspecting a single imputation. It showed an increase in the percent of entries with the event occurring, meaning it would likely provide multiple false positives. Furthermore, the 26 subjects found to have inconsistent diagnoses were removed.

Next, the dataset was modified to the triplet form as described in section 4.1. The column for the time of event/censoring was chosen to be 'M' (months since baseline visit), the 'Event' column was created from the diagnosis, and only the baseline covariates were kept. A general dataset cleanup was also performed by removing redundant and collinear features. Below is a recap of the features removed.

- Baseline measurements were removed due to their redundancy. These include, but are not limited to, CDRSB_{bl}, ADAS11_{bl}, ADAS13_{bl}, EcogPtMem_{bl}, EcogPtLang_{bl}, EcogPtVisspat_{bl}, EcogPtPlan_{bl}, EcogPtOrgan_{bl}, EcogPtDivatt_{bl}, EcogPtTotal_{bl}, EcogSPMem_{bl}, EcogSPLang_{bl}. Please refer to table A.1 for all baseline features.
- Filtered out other redundant columns which did not contribute to the MAR assumption. These were 'PTID', 'RID', 'EXAMDATE', 'IMAGEUID', and 'update_stamp'.
- Removed collinear columns. These are 'Years_{bl}', 'Month_{bl}', 'Month', and 'VISCODE', as they are collinear with the time column 'M' used for the analysis.
- Removed the columns 'PIB', 'DIGITSCOR', and 'FBB' due to too many missing values. They all had over 60% missing entries, which was deemed too severe for imputation given the number of features and size of the dataset.

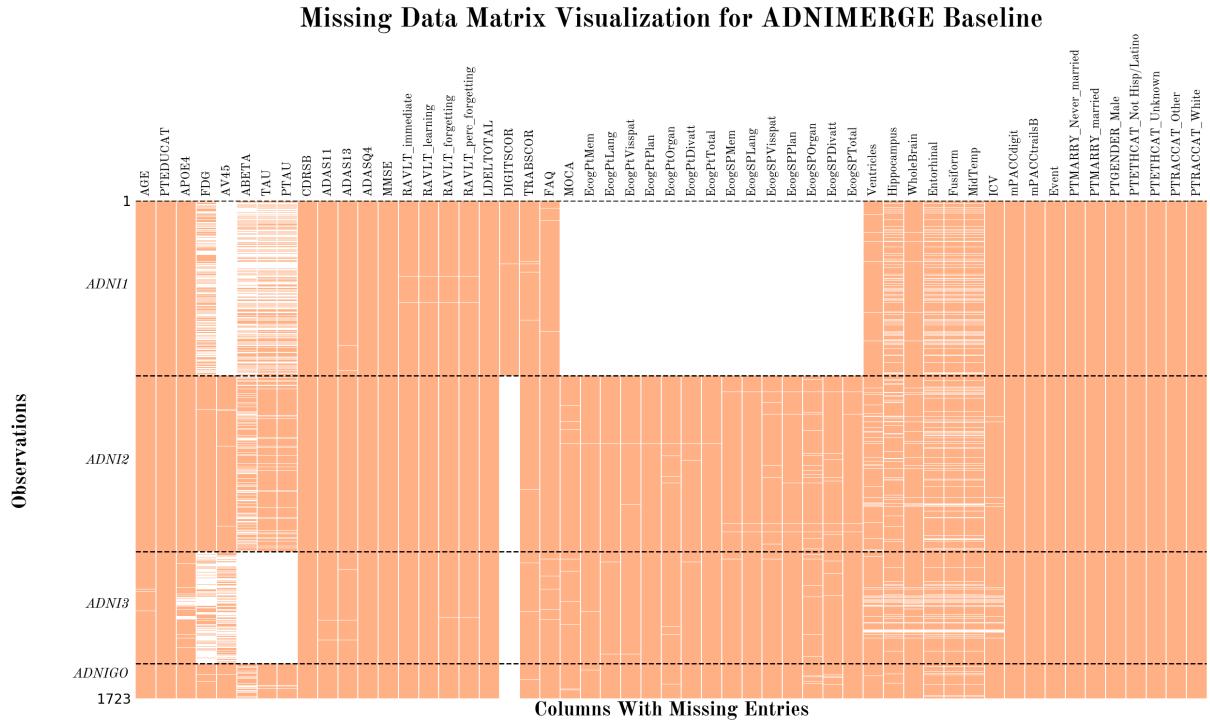


Figure 5.12: Missing data matrix of the baseline measurements from ADNIMERGE.

As neither the machine learning nor the imputation model can handle categorical variables, these had to be one-hot encoded. Due to some features having more than two categories, indicator variables would have to be created. This was done using the `get_dummies()` function in Pandas [102], with the `drop_first` parameter set to true, as to get $k - 1$ dummy variables for k different categories. However, this method may substantially increase the dimensionality of the data. To combat this as much as possible, specific changes were made to categorical values prior, based on the findings of the exploratory data analysis. These are listed below.

- **Martial status:** Reduced to three categories. ‘Divorced’ and ‘Widowed’ were combined into ‘Not Married’, while all ‘Unknown’ values were set to ‘Not Married’.
- **Race:** Reduced to three categories. Categories ‘Asian’, ‘More than one’, ‘Am Indian/Alaskan’, ‘Hawaiian/Other PI’, and ‘Unknown’ were combined into ‘Other’.

Columns ‘COLPROT’, ‘ORIGPROT’, ‘FLDSTRENG’, and ‘FSVERSION’, though unrelated to the machine learning models, may help ensure the data is MAR. Thus, they are kept for the imputation and were also one-hot encoded. The one-hot encoding resulted in 8 new columns. With all this, 66 features were left for the imputation. The dataset was split into a training and test set with an 80/20 split, stratified by the event column. Figure 5.12 shows the missing data matrix of the baseline measurements. There is a lot less missing data for the baseline measurements when compared to figure 5.3.

Finally, the volume measurements of brain regions were normalized by whole brain volume, after which they were normalized to be within [0,1] using Pandas `MinMaxScaler`. This is a common practice in neuroimaging, as the size of regions relative to the whole brain is more important than the absolute size, determined with guidance from CFIN. The final ADNI dataset consisted of 1697 subjects, of which 391 converted to AD.

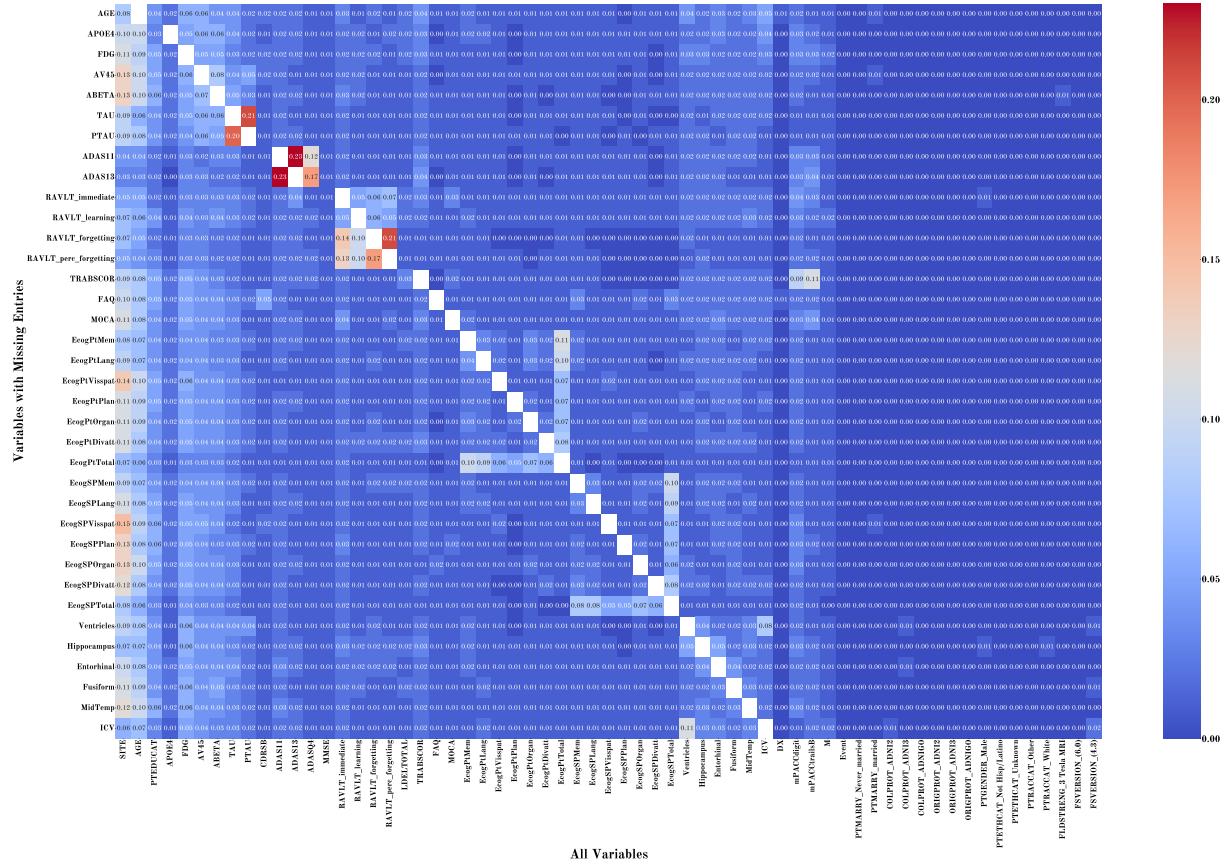


Figure 5.13: Feature importance for data imputation.

Imputation

Using the theory of section 4.2.1, it was decided to impute using MICE. The implementation was done using the Python library miceForest [109], a Python implementation of the MICE algorithm using lightGBM[110].

Following the guidelines of Austin et al. [94] and Van Buuren[96], the number of datasets imputed was set to 20. Ideally, the number of datasets should match the percentage of entries with missing data, which in our case is 100%. However, a limitation in computational power made the time it would take to impute and work with 100 datasets unfeasible. The choice of iterations is less strict and was set to 20 by an iterative process of evaluating the convergence plots. These may be found in the supplementary material. The imputation scheme was PMM imputation with matching subjects identified by the KNN method, and the number of mean match candidates set to 5. The resulting grid of feature importance may be seen in figure 5.13.

It shows some clear paths, such as Ecog features and ADAS features being used to impute one another. After the imputation, some minor cleanups had to be performed. This meant removing the columns that were kept to ensure that the data was MRA.

5.3 Microvascular Biomarkers Dataset

The Microvascular Biomarkers (MVAS) dataset stems from a study from Aarhus University by the Center of Functionally Integrative Neuroscience (CFIN). The goal of the

dataset is to examine the relationship between microvascular biomarkers and the progression of AD. The theory behind the research was covered in section 3.5. Like with ANDI, EDA was performed to inspect the dataset.

5.3.1 Exploratory Data Analysis

MVAS is a combination of two different studies. The first, referred to as MVAS MCI, had the goal of observing conversion to AD in MCI patients. It is a longitudinal study with a follow-up visit after 24 months. The other, MVAS APOE4, studied the microvascular effects of APOE4 in healthy subjects and contains only baseline visits. MVAS APOE4 is ineligible for further analysis as the interest of survival modeling is the longitudinal event. However, it will be used for imputation of missing values. From here on out, 'the dataset' will refer to MVAS MCI unless specifically stated otherwise.

Overview and features

The dataset is rather small. 71 subjects were recruited for an initial baseline observation, of which 57 had a follow-up visit after 24 months. 12 subjects had converted to AD at the follow-up. It contains a mix of demographical, cognitive, and imaging data. The imaging data contains both PET and MRI scans. These cover the following types:

- **PET:** PET scans with binders Pittsburgh compound B (PiB), F-Flortaucipir (FTP), and C-PK11195 were taken. PiB is an $\text{A}\beta$ tracer, and FTP is a tau NFT tracer. C-PK11195 (PK) binds to microglia, which are activated in neuroinflammation [111].
- **sMRI:** Volume scans of brain regions.
- **Perfusion-weighted images (PWI)-MRI:** The PWI-MRI measures of CBF, CBV, CTH, MTT, and RTH measures.
- **Spin echo-weighted images (SEPWI)-MRI:** The SEPWI measures of CBF, CBV, CMRO₂, CTH, MTT, OEF, and RTH.

All the scans were split up into the following regions:

- Frontal, temporal, parietal, and occipital lobes were split into right-hemisphere grey matter, left-hemisphere grey matter, right-hemisphere white matter, and left-hemisphere white matter.
- Hippocampus, thalamus, caudate, and putamen were split into left and right hemispheres.

Whole brain, white matter total, and grey matter total are also stored for each measurement. There are 13 different imaging features, each split into 27 brain regions. With all this, the total number of MVAS features is nearly 400. Considering the number of subjects, the number of features is out of proportion for survival analysis and has to be significantly reduced. A list and description of all the features may be found in table B.1 in the Appendix.

5.3.2 Missing values

Though not nearly as many as ADNI, MVAS does have missing values in the dataset. Figure 5.14 shows the distribution of values for each feature with missing data, both for MVAS MCI and MVAS APOE4.

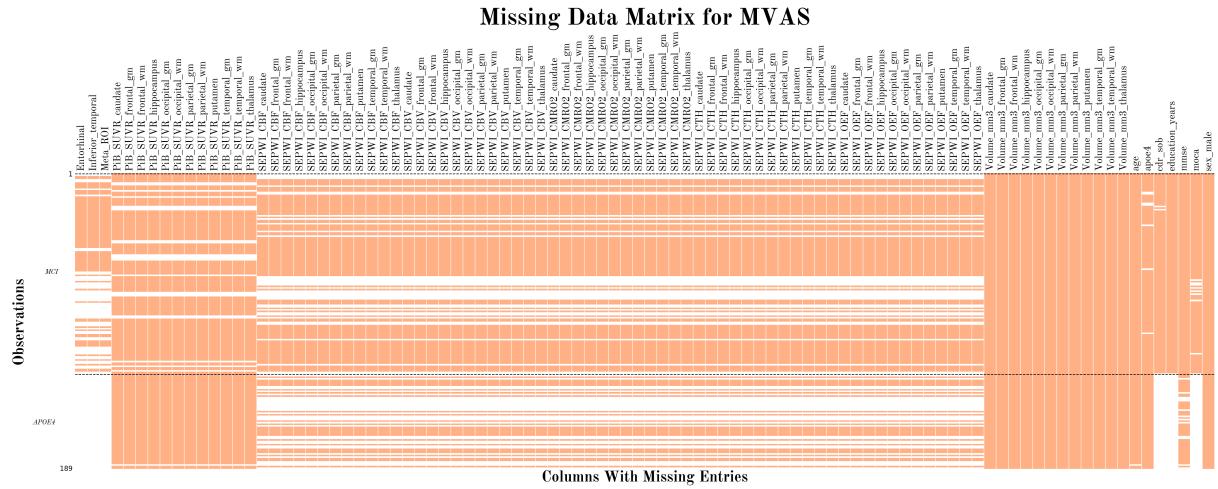


Figure 5.14: Missing entries for the combined MVAS dataset. The dataset shown is from after preprocessing.

Once again, it is assumed that the data is MAR using the same logic as for ADNI. The imputation, however, might be halted due to the small sample size. All subjects have a diagnosis at both baseline and follow-up.

Categorical features

MVAS also has two categorical features. These are plotted in figure 5.15.

Figure 5.15b shows a near-even distribution of zero and one APOE4 alleles, but no subjects with two. Figure 5.15a shows the distribution of years of education, with a big part of subjects having between 11 and 13 years of education.

5.3.3 Data preprocessing

The small size of the dataset and all follow-up measurements at the same time make survival analysis on the dataset very hard. Having all events and censorings happen simultaneously means the baseline hazard can only be estimated for that single time point.

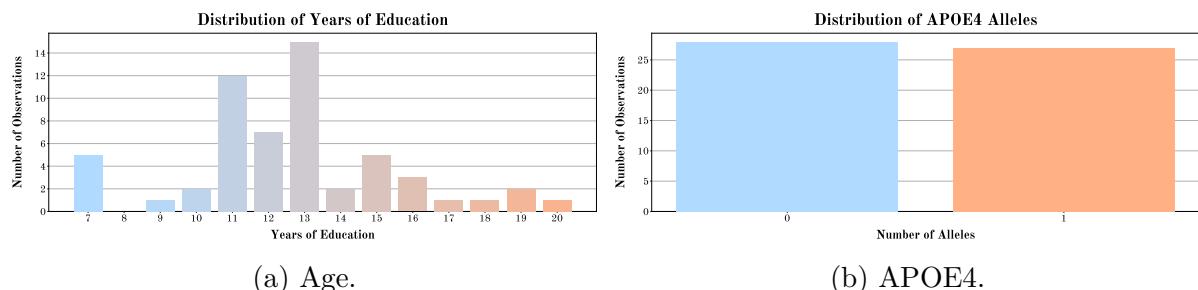


Figure 5.15: Distribution of the categorical variables in MVAS.

As a result, the predicted cumulative hazard function will only provide a prediction for that single time point at 24 months, with no other information on the time horizon. This goes against the argument for using survival analysis. As a result, doing a survival analysis on purely MVAS was deemed to be of no significant interest. Instead, this subsection will focus on preparing MVAS for combining with ADNI.

The first big task of data preprocessing is to reduce the number of features. Spin echo images (SEPWI) are more sensitive to contrast agents in capillary-sized vessels and thus cover the microvascular flow [6]. As this is the variable of interest, the SEPWI measures were kept, except MTT and RTH, as they were derived from other SEPWI features and would otherwise introduce collinearity in the dataset. PWI gives the same measures as SEPWI but for vessels of all sizes. As the goal of this project is to investigate specifically the effect of microvascular disturbances, and not in all vessels, PWI measures were removed. The PK PET measures were also deemed to have too little importance for the goal of this project and thus removed.

To further reduce the dimensionality, the brain regions were modified. The left and right hemispheres were combined for all region-specific measurements, reducing it to 15 regions. Like in the ADNI preprocessing, all volumetric measures were normalized by whole brain volume, after which they were normalized to be within [0,1]. The 'whole_brain', 'gray_matter' and 'nawm' columns were removed as they would otherwise introduce collinearity. This left 12 regions for each imaging measure.

The MVAS MRI dataset is spread across three files: demographical, imaging, and FTP PET scans. Upon request to CFIN, the FTP measures were extracted with an atlas matching ADNI's. Specifically, the measures were extracted for the entorhinal, inferior temporal, and meta-ROI. The three files were merged into one table. The MVAS APOE4 dataset was also split into a demographical and imaging file, which was combined into one table. A column 'COLPROT' was created and set to MVAS_MRI and MVAS_APOE4 for the respective datasets to ensure the APOE4 values would be removed after imputation. The two datasets were then combined into one, as is the same dataset seen in figure 5.14. As the dataset was to be combined with ADNI, imputation would be done at a later timestep. The final MVAS dataset consisted of 57 subjects, of which 12 converted to AD.

5.4 Combining MVAS and ADNI

To create a survival model that utilizes the microvascular features of MVAS, the dataset had to be combined with ADNI. However, the general size difference between the two datasets (57 subjects vs 1697 subjects) would result in a very high missing percentage for the unique MVAS features upon a vertical merge. To combat this, it had to be combined with a single cohort of ADNI. The criteria for choosing the subset would be to have as many overlapping features as possible to optimize PMM. ADNI1 is the most widely used cohort in related works and is the cohort with most events. However, it doesn't have MoCA test scores, nor does it have any FTP measures. Even though it has PiB PET scans included, only an insignificant amount of subjects had them at baseline. ADNI GO, though it had MoCA scores, suffered the same issues. Furthermore, it only had follow-ups for 24 months, essentially. Though ADNI2 has some FTP measures, none were at baseline, and only 87 observations had the measurement. ADNI3, on the other hand, has both MoCA and FTP measures. A downside to ADNI3 is the number of events, as

only 47 subjects converted to AD. Despite this, ADNI3 was chosen to be combined with MVAS due to the number of overlapping features.

Subjects overlapping from previous cohorts had their time measurement 'M' shifted to start at 0. Entries with missing diagnoses were removed, leaving 628 subjects and 3619 entries. Only a selection of features was kept. These were overlapping demographical variables (age, education, gender) and cognitive scores found to be of high importance in related works (CDRSB, ADAS11, ADAS13, ADASQ4, MMSE, RAVLT immediate, LDELTOTAL, FAQ, MOCA). Biomarkers kept were APOE4 and FTP PET measures. An attempt to include volumetric measurement from ADNI was made also made, but due to atlas differences between MVAS and ADNI, this was infeasible.

This left 16 features, of which ten overlapped with MVAs. The size of ADNI3 was still too large relative to MVAS, meaning it had to be further reduced. Thus, it was decided to only use MCI subjects. This left 253 subjects, of which 44 events were observed. The same would have to be done for MVAS.

In order to combine the overlapping columns, MVAS columns were renamed to match ADNI. MVAS, and ADNI were vertically combined. Both datasets have features that aren't present in the other, meaning these would become empty variables that had to be imputed. The combined dataset, like ADNI, was transformed into the triplet form. It was then split into training and test sets (80% and 20%) stratified by the event column. Thus, only baseline measurements were kept. Now, two different imputations were performed:

- **MVAS-ADNI-A (MA-A):** Baseline measures from ADNI3 MCI subjects and all MVAS MCI measurements eligible for survival analysis in were used for imputation.
- **MVAS-ADNI-B (MA-B):** In addition to MVAS-ADNI-A, all additional data from MVAS was added to the training set. This accounted for HC subjects (also from APOE4), and subjects with only baseline measurements. This increased the train set by 113 measurements, which would be removed after imputation.

MA-A helps avoid potential bias induced by introducing non-MCI subjects into the imputation. MA-B was done as an attempt to reduce the comparable size difference and provide more measures for the imputation. However, as the subjects are non-MCI, these could potentially skew the imputation. Both have their pros and cons and would have to be tested for a better evaluation. An overview of the complete final dataset before the train/test split, including extra MVAS entries from MVAS-ADNI-B, may be seen in figure 5.16. The final dataset consisted of 291 subjects, 56 of which converted to AD. To compare the effect of MVAS, pure ADNI3 with equivalent preprocessing was imputed for testing. This dataset consisted of 253 subjects, of which 44 converted to AD.

Considerations were made if first imputing ADNI3 and MVAS separately would be doable. However, imputing M_{adni} and M_{mvias} datasets for each, and then imputing $M_{MA} = M_{adni} \cdot M_{mvias}$ datasets was deemed computationally infeasible.

Imputation for ADNI3, MA-A, and MA-B were all performed with the same parameters. Twenty-five datasets were imputed over 20 iterations. It was ensured that MVAS subjects in the test set would not have their extra measurements included in the training imputation. Once again, MICE with PMM was used for the imputation. After imputation, columns that helped ensure MAR such as 'COLPROT', 'DX' and 'RID' were removed.

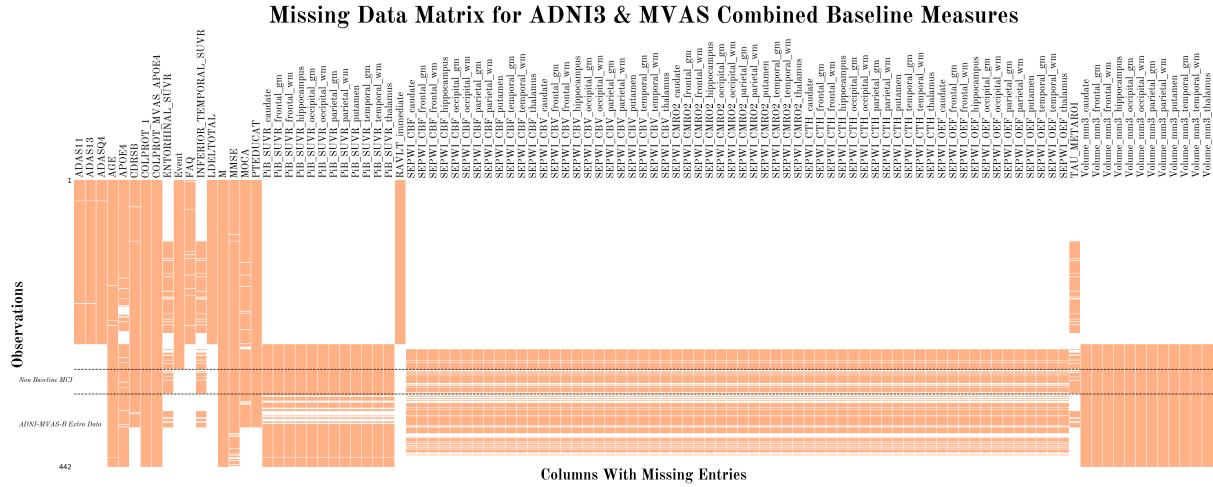


Figure 5.16: Missing value distribution of MA-B.

5.5 Survival Modeling

The choice of model for survival analysis is non-trivial as it is based on the criteria defined by the dataset and the problem statement. The most important aspect of this project is interpretability; we wish to understand the covariates' effect on the prediction. Another criterion is for the model to handle high dimensional data, as the datasets have many features compared to events. To incorporate the strengths of MI, the models should preferably be poolable following the guidelines from section 4.2.1. Finally, multiple models were chosen to obtain a clearer picture of performance.

All these criteria are met by the penalized Cox PH models (Lasso, Ridge, and ElasticNet) introduced in section 4.1. The obtained coefficients show a direct effect of each covariate. The coefficients may be pooled by following formulas 4.10 through 4.13 to obtain a final model. Finally, the penalty term introduced should help combat overfitting and make the model handle high dimensional data, which was also shown by Spooner et al. [20]. Recalling the findings from chapter 2, RSF was found to obtain high estimation scores and was shown to generalize well to high dimensional data across multiple studies. The black-box nature of the model may be handled using the permutation importance technique. A clear downside, though, is the inability to pool the models.

These are all rather traditional approaches. Deep learning-based models, such as Deep-Surv, were considered. Their lack of interpretability and their tendency to overfit, combined with the rather small datasets, made them less desirable. Pooling the models would also be infeasible.

For these reasons, the models considered for implementation were Cox PH Ridge, Cox PH Lasso, ElasticNet, and RSF. A Cox PH model without any penalty term would also be included as a baseline measure.

5.5.1 Implementation

The survival models were implemented using lifelines[105] and scikit-survival[104].

Cox Proportional Hazard Models

Both lifelines and scikit-survival have implementations for all the Cox PH models. However, both had a downside: lifelines did not allow the manual insertion of coefficients into the model after pooling, which prevented testing. scikit-survival, on the other hand, does not provide within variance for the coefficients, making pooling by Rubin's rules impossible. A rather 'hacky' workaround was to use a hybrid of the two: use lifelines for creating the initial models and pooling the coefficients and then afterward set the coefficients in a scikit-survival implemented model, by which the model could be tested.

The penalty term was found using a 5-fold cross-validation grid search with the C-index as the scoring parameter. The coefficients were pooled using Rubin's rules. However, the baseline hazards were only averaged for the final model, as no variance was available.

Random Survival Forest

The RSF was implemented using the scikit-survival implementation. The model's hyperparameters were also found using a 5-fold cross-validation grid search using the C-index as the parameter. The tunable parameters were `n_estimators`, `min_samples_split`, `min_samples_leaf`, and `max_features`. The model was then trained on the entire training using the best-found hyperparameters.

CHAPTER 6

Experiments and Results

6.1 Experiments

Upon preprocessing ADNI and MVAS, five different datasets were obtained. These may be split into two main groups. The cleaned ADNIMERGE was used as a complete dataset (ADNI-A), while it was also split into only subjects with MCI diagnosis at baseline (ADNI-M). These are referred to as the ADNIMERGE-based datasets. The remaining datasets, ADNI3, MA-A, and MA-B, are referred to as the MVAS-based datasets. Experiments were conducted on each of the datasets using the previously described models. The two groups were tested with different goals in mind.

The ADNIMERGE-based datasets were tested with performance optimization in mind. The MVAS-based datasets were instead tested to examine the effects of the microvascular biomarkers, acting as a proof-of-concept of their effectiveness. With this, interpretability and understanding of their effects was the primary goal. Table 6.1 presents an overview of all datasets split into the train and test sets. As MA-A and MA-B have the same subjects, both are covered in the same column.

	ADNIMERGE (All)	ADNIMERGE (MCI)	ADNI3	MVAS+ADNI3
Subjects	1357/340	764/202	202/51	232/59
Event	313/78 (23%)	285/73 (36%)	35/9 (15%)	45/11(19%)
Features	50	50	16	99

Table 6.1: Number of subjects, events and features in the training and test sets of each method.

Evidently, the datasets vary greatly in size. This means the hyperparameter grid would have to be set accordingly. For the ADNIMERGE-based datasets, the hyperparameter grid for the RFS model was:

- n_estimators: [200, 400, 600, 800]
- min_samples_split: [3, 5, 10]
- min_samples_leaf: [3, 5, 10, 15]
- max_features: [sqrt, log2]

It allowed for larger trees as the datasets were larger and, in turn, less prone to overfitting. A specific hyperparameter search was performed for both ADNIMERGE-based datasets. The datasets are quite large, so the cross-validation takes a while. Because of

this, cross-validation was only performed on one of the imputed datasets and reused on all the other. The grid used for hyperparameter search for MVAS-based datasets was:

- n_estimators: [100, 200, 300, 400]
- min_samples_split: [4, 6, 8, 10]
- min_samples_leaf: [4, 6, 8, 10]
- max_features: [sqrt, log2]

The number of trees is lower with a smaller range of possible split values. Hyperparameter search was performed on all imputed datasets for both ADNI3, MA-A and MA-B.

The penalized Cox models had different penalties tested for MVAS-based- and ADNIMERGE-based datasets. MVAS-based datasets had higher penalties to combat the small size of the datasets in an attempt to generalize better to the many features. The following penalties were tested with cross-validation for all models:

- MVAS-based penalty: [0.2, 0.3, 0.4, 0.5, 0.6, 0.7]
- ADNIMERGE-based penalty: [0.01, 0.05, 0.1, 0.2, 0.3, 0.4]

ElasticNet had the α value set to 0.5 to allow for both L1 and L2 penalties equally. After iterative testing, the Lasso regression was reduced to [0.05, 0.1, 0.2, 0.3] for the MVAS-based datasets, as otherwise, all but one coefficient would be set to 0.

The pooled Cox models would all be evaluated on each of the M test sets for each respective dataset. The final score would be an average, including a standard deviation. As the survival forests couldn't be pooled, an average of each model's score on each test would be averaged. Permutation importance was performed for the future importance, the result of which would then be pooled using Rubin's rules. The metrics used for testing were C-index, IPWC-C statistic, and IBS, where the goal of using a wider range of evaluation metrics would be to obtain a more diverse perspective of performance.

6.2 Results

This section will present the results from the experiments, which cover the score of the models and investigative plots into their performance. Only a subset of all plots are shown in this section. For all plots, please refer to the supplementary materials. The results are split into two parts. The results of the ADNIMERGE datasets are presented first, followed by the results of the MVAS-based datasets.

6.2.1 ADNI MERGE

All scores from all models on the ADNIMERGE-based datasets are shown in table 6.2. ADNI-A obtains a high C-index across all models. Each has a score above 0.9, indicating they rank 9/10 pairs correctly. Remembering the literature of section 4.1, as only 23% of subjects experienced the event, the C-index might be biased upwards. This is reflected in

both the IPCW-C statistic and the IBS. The Cox models with ridge and Lasso penalties seem to perform the best considering all scores, though not by much. The RSF model also seems to perform well with a high C-index and a low IBS.

The results of the ADNI-M further imply that the ADNI-A C-index scores are biased upwards. The dataset performs worse in the C-index, with all models having scores within 0.84 to 0.86. The IPCW-C statistics are about even across both datasets. The ElasticNet and Lasso models seemed to perform best based on all scores.

Table 6.2: Mean IPCW-C, IBS, and C-Index with standard deviation on the M imputed versions of each dataset. The Cox-based models were tested with pooled coefficients, while RSF was an average of the M models' performance.

Model	Metric	ADNI-MERGE (All)	ADNI-MERGE (MCI)
Cph	C-Index	0.901 ± 0.003	0.841 ± 0.005
	IPCW-C	0.837 ± 0.009	0.843 ± 0.007
	IBS	0.121 ± 0.003	0.149 ± 0.149
CphRidge	C-Index	0.920 ± 0.003	0.860 ± 0.005
	IPCW-C	0.837 ± 0.008	0.844 ± 0.006
	IBS	0.142 ± 0.002	0.152 ± 0.003
CphLasso	C-Index	0.918 ± 0.003	0.857 ± 0.006
	IPCW-C	0.838 ± 0.009	0.847 ± 0.011
	IBS	0.156 ± 0.003	0.120 ± 0.003
CphElastic	C-Index	0.918 ± 0.003	0.857 ± 0.006
	IPCW-C	0.836 ± 0.009	0.846 ± 0.009
	IBS	0.173 ± 0.002	0.116 ± 0.003
RSF	C-Index	0.907 ± 0.004	0.841 ± 0.007
	IPCW-C	0.854 ± 0.009	0.832 ± 0.010
	IBS	0.121 ± 0.003	0.149 ± 0.004

ADNI-A inspection

The results of the ADNI-A models would have to be investigated further. It was suspected that the model utilized the skewed event distribution and would be too optimistic, resulting in the biased C-index. For this, the ridge model was inspected closer. Figure 6.1 shows the distribution of the coefficients for the ridge model.

The figure shows that many of the features considered of high importance in the literature are also weighted highly in the model, such as lower volume in the entorhinal and hippocampus regions or increased AV45, a tracer for A β . Other features, such as the subject's age, were weighted low, which goes against the literature. These were general trends seen across the Cox models. The RSF permutation results showed very small values for every coefficient, with relatively large confidence intervals.

Investigating the model's prediction, figure 6.2a shows the distribution of the true survival time subtracted from the predicted survival time (marked when the survival function is under 0.5) for all subjects that experienced the event.

It essentially confirms the suspicion that the model is overly optimistic. It shows that the model always predicts a survival time that is too high. This is true for all the Cox

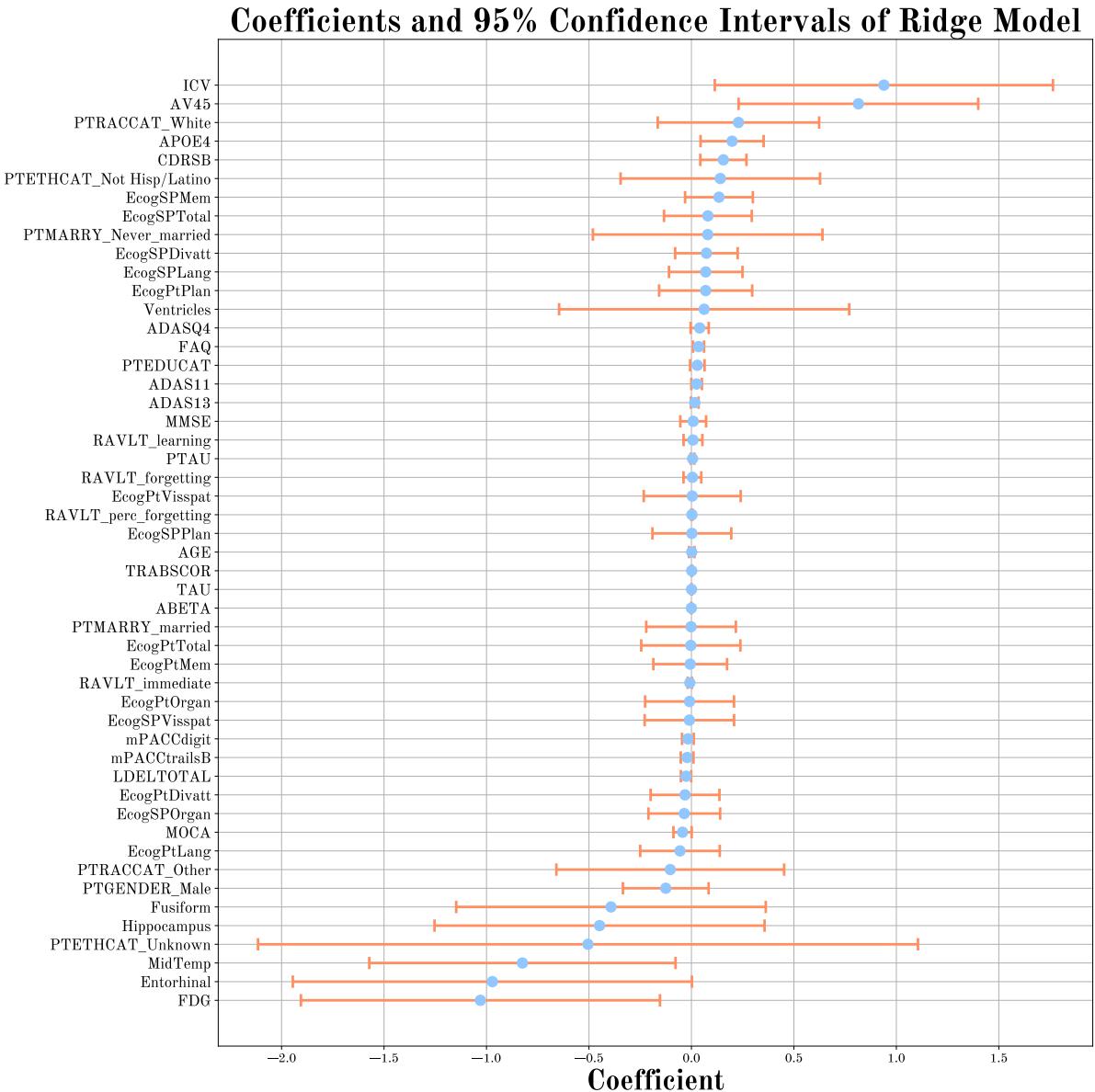
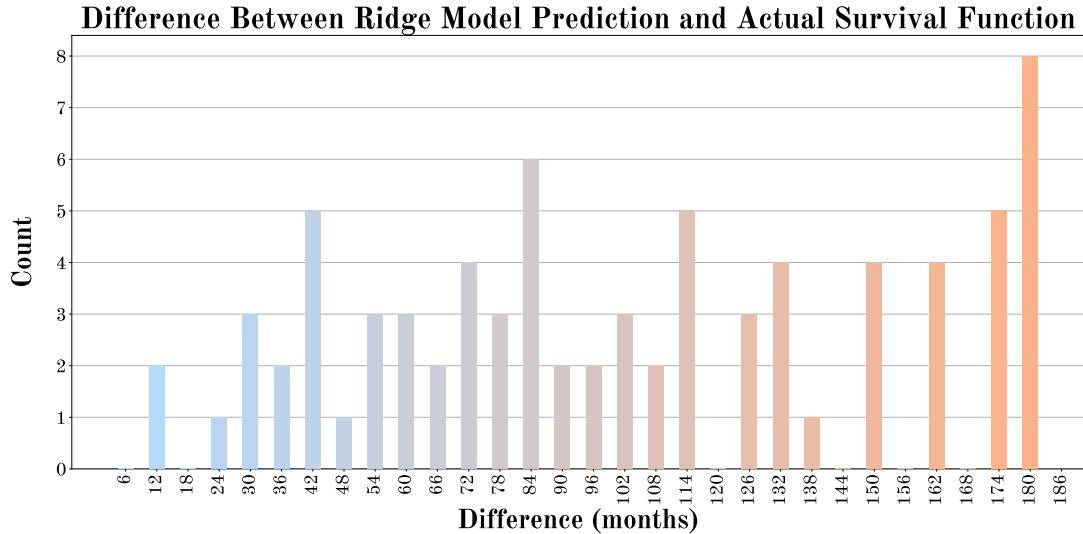


Figure 6.1: ADNI-A trained ridge model. The spread of the pooled coefficients with 95% confidence interval is plotted.

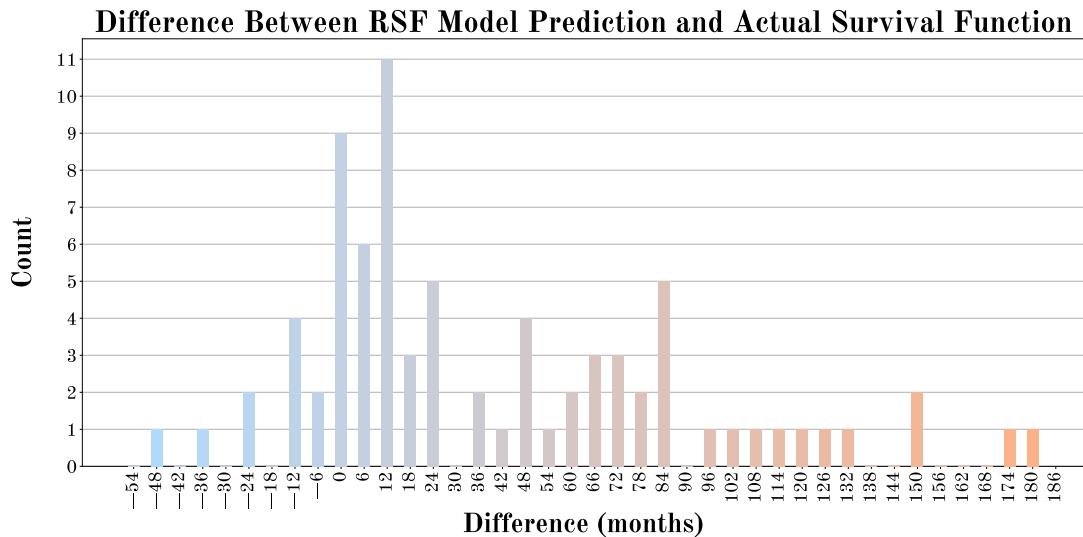
models, which may be seen in the supplementary plots. The RFS model was also plotted (by choosing one of the models), showing a better prediction ability on the data, as shown in figure 6.2b. It shows a more even spread around 0, indicating better predictive performance.

ANDI-M inspection

Models trained on ADNI-M were hypothesized to perform better, as the event distribution was more even. Furthermore, the IPWC-C statistic and C-index were more in line with the C-index. As the best-performing model, ElasticNet was inspected further. Figure 6.3 shows the difference distribution between predicted and true survival time for all subjects that experienced the event.



(a) ADNI-A trained ridge model.



(b) ADNI-A trained RSF model.

Figure 6.2: Predictions of the ADNI-A models. Shows the true survival time subtracted from the predicted survival time for all subjects that experienced the event.

The figure shows that the model performs much better than the ADNI-A models. A majority of the subjects were predicted within 12 months of the true survival time, though outliers were also present. This aligns with the C-index and the IPWC-C statistic and meets the expected performance. This goes across all the penalized models, whereas the RSF model from inspection performs overly optimistic.

6.2.2 MVAS-ADNI

All scores for the MVAS-based datasets are shown in table 6.3. Cox models without penalty terms were unable to be trained due to the high collinearity between features for the MA-A and MA-B datasets. The C-index for ADNI3 is generally high, with the Ridge penalized model obtaining a score of 0.911. The IPCW-C statistics are a bit worse, though they are also relatively high. However, the IBS is quite high for all the penalized models,

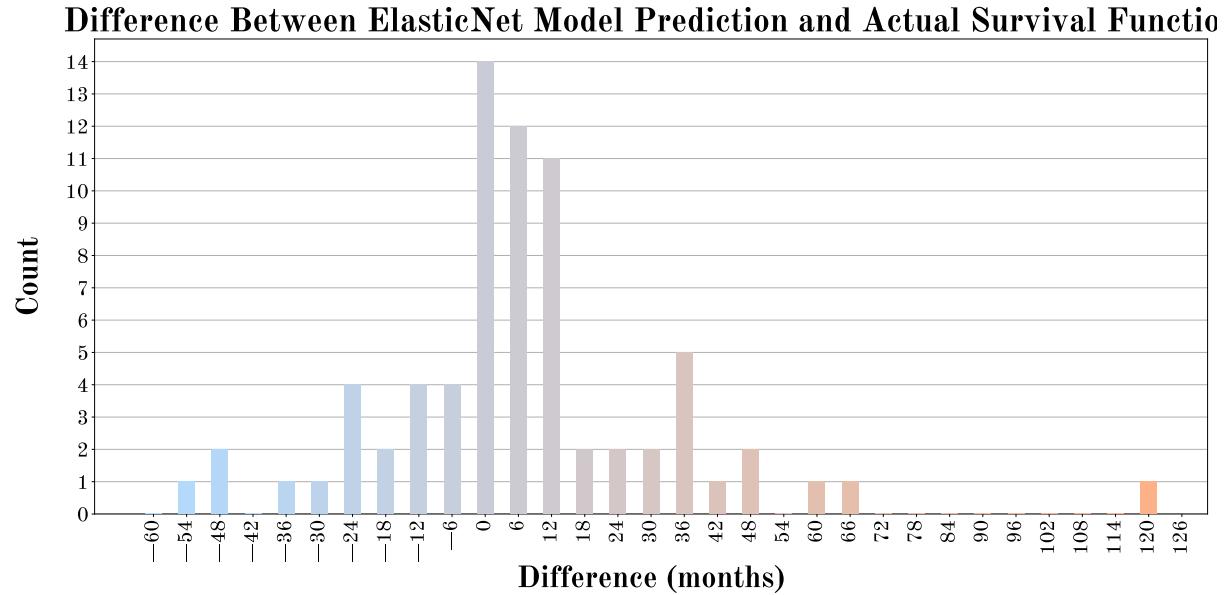


Figure 6.3: ADNI-M trained ElasticNet model difference between predicted and true survival time.

indicating that they might not perform as well as the C statistics suggest. Interestingly, the Cox model without penalty performed well based on the IBS score, surpassing all penalized models and matching the RSF model. Based on all metrics, the RSF model performed the best.

Table 6.3: Mean C-Index, IPCW-C, and IBS with standard deviation on the M imputed versions of each dataset. The Cox-based models were tested with pooled coefficients, while RSF was an average of the M models' performance.

Model	Metric	ADNI3	MVAS+ADNI-A	MVAS+ADNI-B
Cph	C-Index	0.878 ± 0.013	-	-
	IPCW-C	0.890 ± 0.015	-	-
	IBS	0.108 ± 0.108	-	-
CphRidge	C-Index	0.911 ± 0.010	0.735 ± 0.035	0.715 ± 0.052
	IPCW-C	0.906 ± 0.029	0.851 ± 0.032	0.840 ± 0.042
	IBS	0.231 ± 0.001	0.230 ± 0.002	0.656 ± 0.015
CphLasso	C-Index	0.894 ± 0.017	0.736 ± 0.040	0.706 ± 0.052
	IPCW-C	0.873 ± 0.050	0.855 ± 0.034	0.833 ± 0.037
	IBS	0.235 ± 0.235	0.240 ± 0.000	0.196 ± 0.003
CphElastic	C-Index	0.899 ± 0.018	0.730 ± 0.043	0.698 ± 0.051
	IPCW-C	0.884 ± 0.049	0.847 ± 0.032	0.829 ± 0.035
	IBS	0.235 ± 0.235	0.238 ± 0.000	0.233 ± 0.000
RSF	C-Index	0.904 ± 0.014	0.761 ± 0.040	0.754 ± 0.052
	IPCW-C	0.892 ± 0.046	0.851 ± 0.040	0.844 ± 0.062
	IBS	0.108 ± 0.011	0.132 ± 0.008	0.134 ± 0.010

Of the MA-A and MA-B models, the MA-A performed the best across nearly all parameters. This indicated that including the HC subjects in the imputation was not

beneficial. The scores of the models were lower than ADNI3 across all parameters. The highest C-index was 0.761 for the RSF model, which also performed the best across all parameters. The IPCW-C statistic is much higher than the C-index for all models, indicating that the model's ability to compare individuals is better than the C-index suggests. However, it is unclear how much importance should be put into this. A further inspection of the models now follows.

MA-A

Though the RSF model performed the best based on test scores, a further inspection of the Cox models was needed to help answer why. The pooled coefficients for the Ridge model are shown in figure 6.4. Most features had large confidence intervals, which was to be expected considering the missing data percentage, the number of features, the fact that it was a small dataset, and the low number of events.

Decrease in frontal white matter SEPWI features CBF, OEF had rather high absolute coefficients. The capillary dysfunction hypothesis of section 3.5 postulated a decrease in OEF, by which negative coefficients in the Ridge model would align with the hypothesis. However, the other models have near-zero coefficients for all OEF measurements. A notable observation is that the CTH measurements in all regions for all models have near-zero coefficients and similarly small confidence intervals. This goes against the findings of the original MVAS study, which found significant differences in CTH between pAD-MCI and SNAP-MCI. Across all three models, CBF frontal had negative coefficients, following the study findings that CBF decreases in AD. Some of the models had CBV measurements with relatively high coefficients, which in the study found no significant change over time. However, CBV showed to have both negative and positive coefficients in different regions, which might indicate that it is simply a result of the large variance.

Generally, other features than the microvascular changes showed higher predictive performance. Volume measurements, FTP PET measurements, and $CMRO_2$ (a metabolic rate measurement) generally had higher absolute coefficients. Though the cognitive scores were all relatively low in the Ridge model, they were of relatively high importance in the ElasticNet, Lasso, and RSF models. The FTP measurements score was also high across all models, possibly due to their initial low missingness.

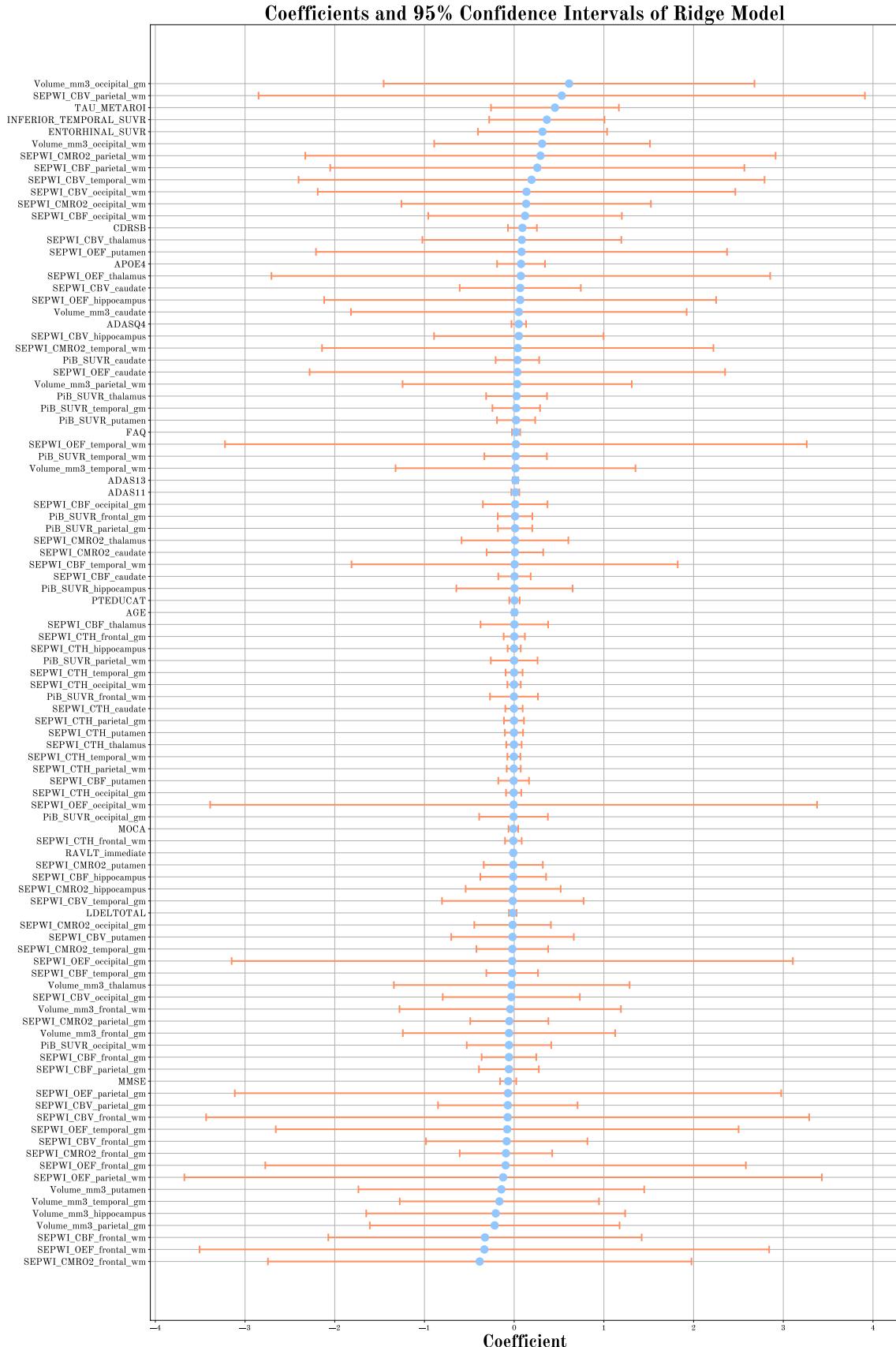


Figure 6.4: Pooled coefficients for the Cox model with ridge penalty on the MA-A dataset, along with 95% confidence interval.

To further inspect the performance of the models, figure 6.5 plotted the difference in true survival time and predicted survival time. It gives a clear indication as to why the IBS scores were low. It seems that the model is overly optimistic and never predicts that the event will occur.

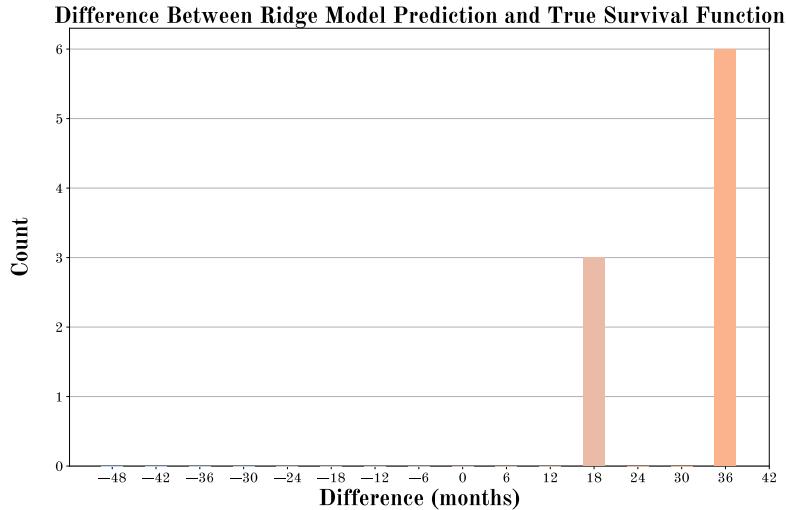


Figure 6.5: MA-A trained Ridge model distribution of the difference between predicted and true survival time.

A likely cause of this is the low number of events in the dataset in combination with the proportional hazard form. It requires a certain number of events to create a suitable baseline hazard. The baseline hazards of the penalized models are shown in table 6.4. The table shows they are minimal, and at multiple time points even 0. The resulting predicted hazard function is very low, predicting no subject will experience the event. This is likely also why the RSF models performed better, as the underlying structure differs.

Table 6.4: MA-A dataset baseline hazards for Ridge, Lasso, and ElasticNet models. Time-points are in months.

Timepoints	M12	M18	M24	M30	M36	M42	M48	M54	M60	M66
Ridge	0.033	0.004	0.076	0.029	0.034	0.000	0.089	0.000	0.135	0.000
Lasso	-	0.006	0.104	0.033	0.037	0.000	0.093	0.000	0.137	0.000
ElasticNet	-	0.006	0.100	0.035	0.039	0.000	0.097	0.000	0.145	0.000

Instead, more emphasis should be put on the predicted risk score, which is reflected in the C-index and the IPCW-C statistic. Figure 6.6a shows the distributions of the predicted risk scores, comparing the ones for censored subjects to subjects with events. It shows a clear distinction between the two groups, indicating that the model is able to separate them rather well. However, there are outliers, which is also reflected in the C-index.

To see the prediction of the RSF model, figure 6.7 shows the distribution of the difference between predicted and actual survival time. It shows that the model models fit better than the Cox models, as some events are predicted to occur, once again reflected by the IBS result.

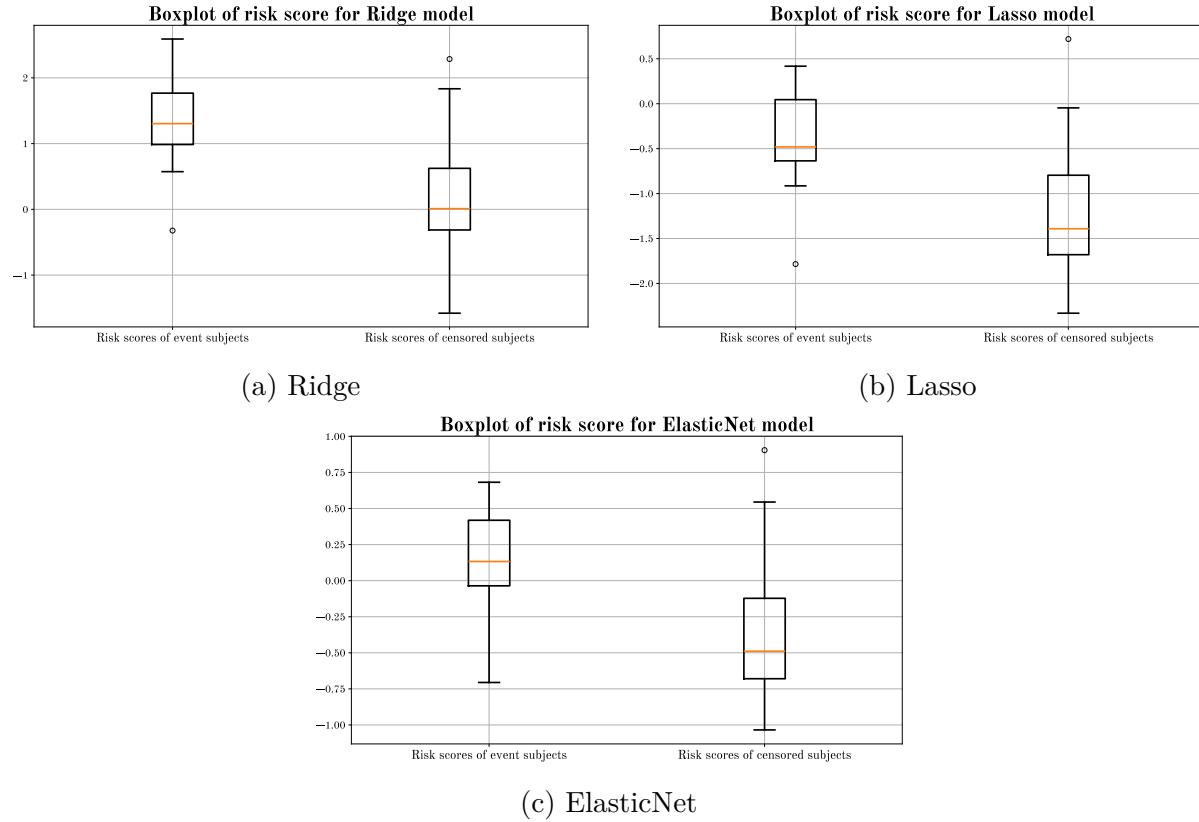


Figure 6.6: MA-A trained models' predicted risk scores on censored subjects and subjects with events.

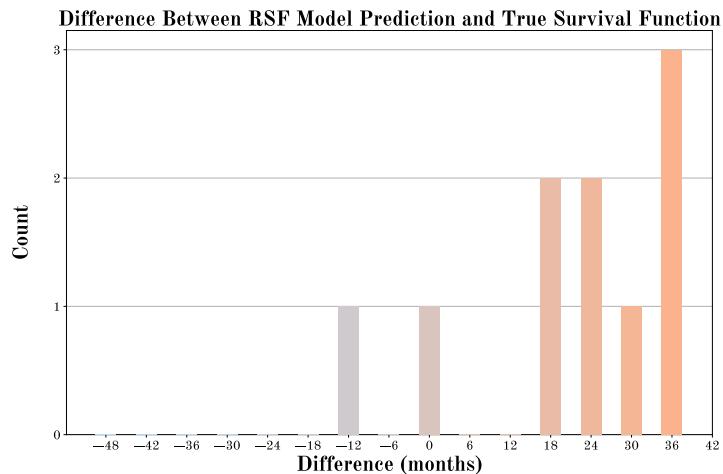


Figure 6.7: MA-A trained RSF model distribution of the difference between predicted and true survival time.

ADNI3

Finally, the results of the ADNI3 models are inspected. The pooled coefficients for the Ridge model are shown in figure 6.8. Though the confidence intervals are large, the FTP PET features have the largest features, which is also true for the other Cox models.

Inspecting the model's prediction, figure 6.9a shows the difference between the pre-

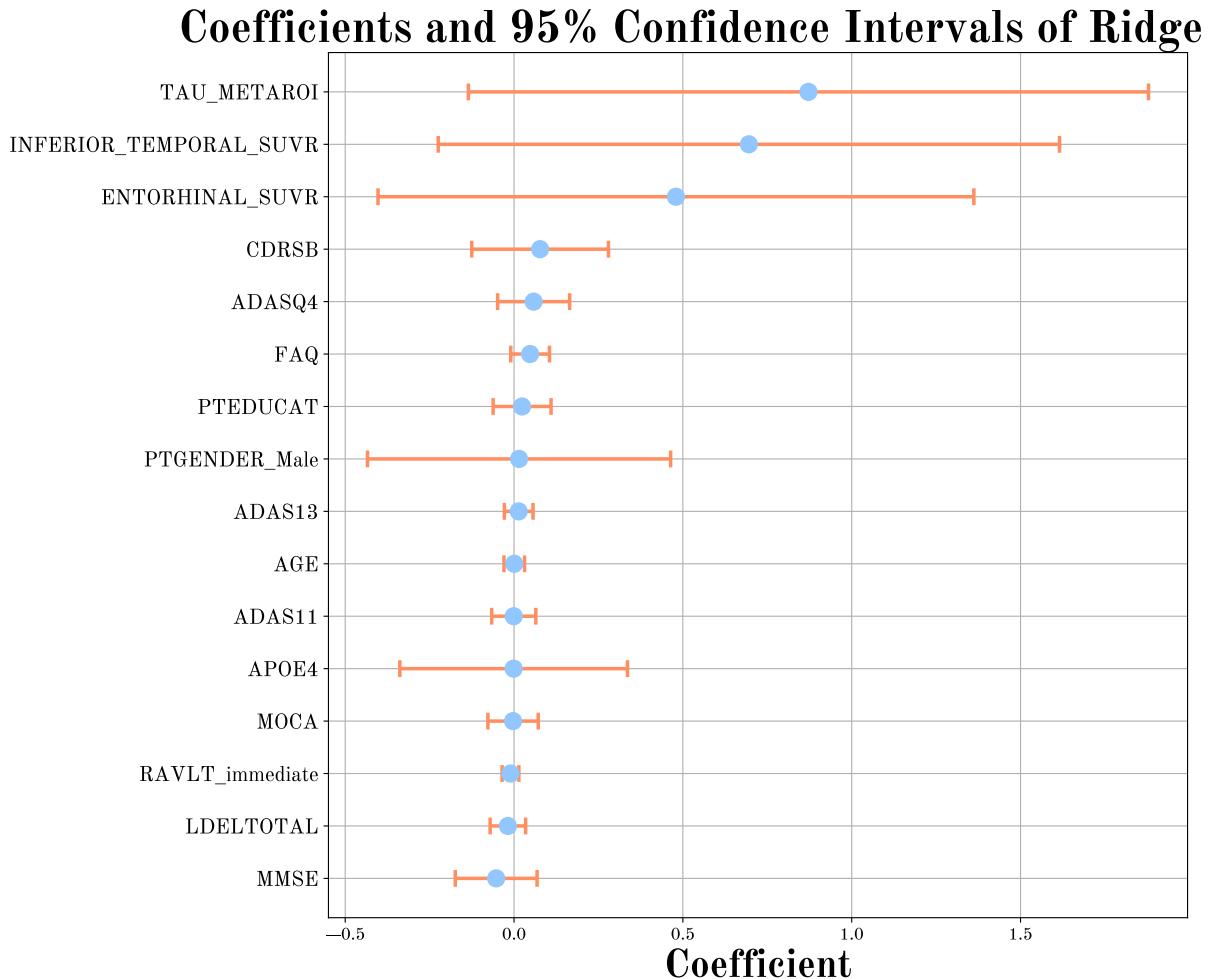


Figure 6.8: ADNI3 trained Ridge model distribution coefficients with 95% confidence interval.

dicted and true survival time for the Ridge model. The model does not predict any event, as reflected in the IBS score. Like with the MA-A models, it suffers from low baseline hazards, though not to the same extent. In comparison, the RSF model performs better, as shown in figure 6.9b. It shows a more even spread around 0, indicating better predictive performance. This highlights that the problems in the MA-A dataset stem from the ADNI3 dataset.

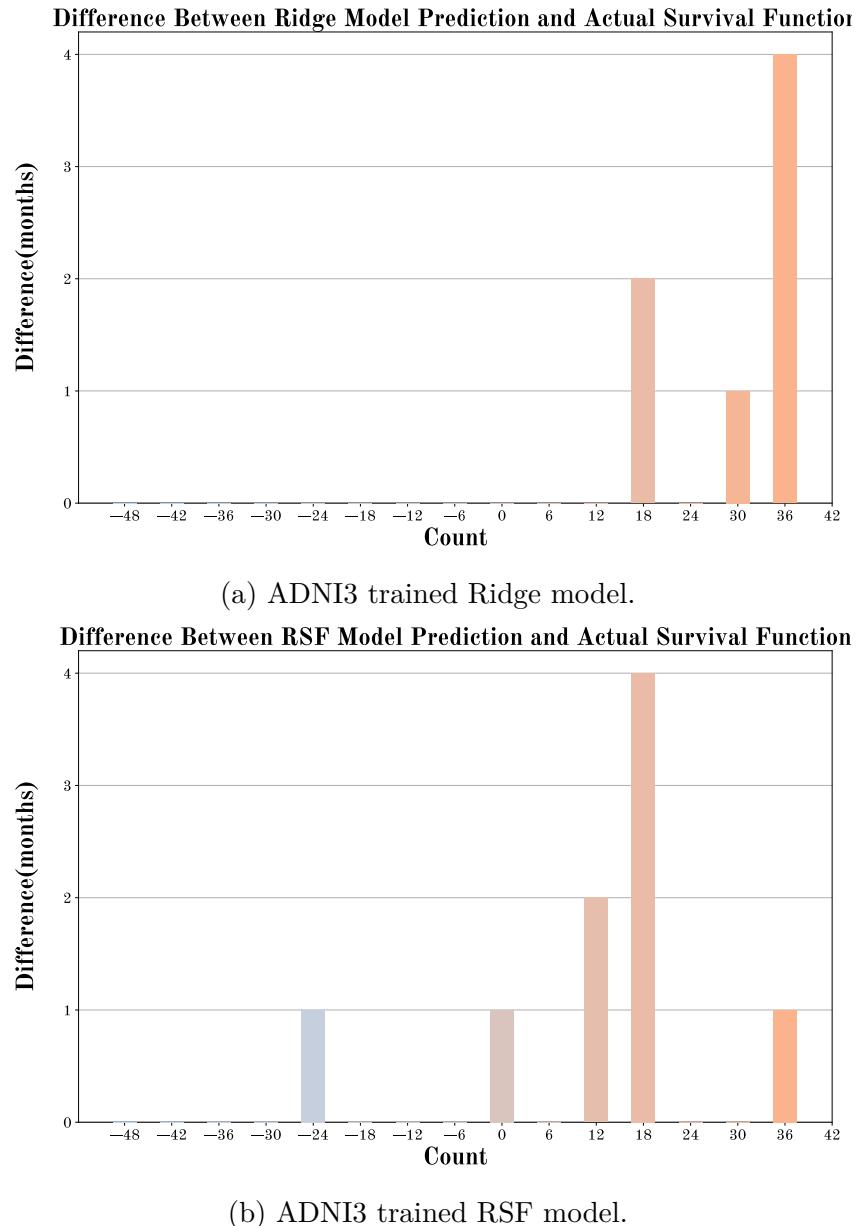


Figure 6.9: ADNI3 models distribution of the difference between predicted and true survival time.

Discussion, Conclusion & Future Work

7.1 Discussion

Involving technology in AD prediction is an increasingly important task as the number of cases continues to rise alongside the increasing average age of the population. Incorporating time-to-event analysis is a natural step in this direction, though it is not nearly as widespread as traditional machine learning. Furthermore, the newfound effect on the brain microvasculature has yet to be applied to survival analysis. This thesis found that very few had performed survival analysis using multimodal data on the entire ADNI dataset, which it set out to do. Furthermore, it explored the MVAS dataset and explored its combination with ADNI, aiming to assess the effects of its unique microvascular features. The heterogeneous nature of clinical data makes handling missing variables an essential part of preprocessing the data. Upon cleaning the datasets, MICE was employed with a PMM scheme to impute the missing entries, after which survival analysis on each dataset was done using traditional statistical and machine learning survival models.

The imputation of missing values is a crucial aspect, yet there is no single best imputation model, as testing for how well the missing values are imputed is impossible as they are just that: missing. The use of MICE helped model the uncertainty of the final Cox models. Furthermore, the use of PMM helped keep the distribution of the original data. Ideally, the imputed number of datasets would be higher, though computational limitations made this impossible. This was especially true for the MVAS-based datasets, as the missingness was very high, resulting in a correspondingly high variance.

Various metrics were used to determine the performance of the models in the form of C-index, IPWC-C, and IBS. The ADNIMERGED-based datasets showed varying results depending on what set of the dataset was used. Using the entire dataset showed that the skew in events was not handled well by the Cox methods, resulting in overly optimistic models. This was not apparent through the C-index, which was likely biased upwards. Instead, it was implicated by the IPWC-C statistic and IBS. The RSF model, on the other hand, generalized the data better, proving to be a rather powerful method, as discovered in the state-of-the-art research. Machine learning models without the proportional hazard form seem better suited for survival modeling on the entire ADNI dataset, whereas Cox models don't.

Using only patients with an MCI diagnosis resulted in models with a lower C-index, but better actual predictive ability based on visual inspection. The ElasticNet model achieved a C-index of 0.857, near state-of-the-art performance on the complete ADNI dataset with MCI to AD conversion as the event. It shows the potential of using multimodal data on

the entire dataset and would perhaps perform even better with more complex models.

The use of the MVAS dataset did not provide the results that were hoped for. The inability to train a model purely on MVAS due to the size and times of observations made it hard to evaluate the unique biomarkers. Furthermore, the uncertainty introduced by combining ADNI and MVAS and introducing huge percentages of missing values made the results too have high variance and lack a general structure. Increasing the number of datasets imputed could be a way to combat this, though it requires more computational resources than was accessible for this project. What could be observed was that many of the more widespread biomarkers, such as FTP PET and cerebral atrophy, as well as the cognitive scores, outperformed the microvascular biomarkers in general. One may argue that training purely with the microvascular biomarkers would have made more sense. However, including other biomarkers directly compared their predictive performance to other features, showing that they did not perform as well.

The flaws of ADNI3 as a standalone dataset for Cox models were also highlighted. The low number of events and the short follow-up time meant the baseline hazards were too small to have any predictive power. This problem persisted when combined with MVAS, making survival prediction impossible. It is worth noting that this might have affected the microvascular biomarkers' poor results.

The risk score of AM-A and ADNI3 could still be used as an evaluation despite the baseline hazards, which showed promise when comparing the risk score of subjects with events and censored subjects and obtaining a C-index of around 0.73 for all Cox models.

An important point of discussion is the diagnoses in the datasets. As mentioned in the theory, the diagnosis of AD is not straightforward and is often based on cognitive scores. This essentially introduces a bias towards cognitive scores as predictors. Human error in diagnosis will also introduce bias to the model if not handled properly, as seen in the cases of patients diagnosed with AD regressing to MCI. Furthermore, the amount of missing entries in the diagnosis in ADNI is substantial, possibly introducing interval censoring in the worst case.

7.2 Conclusion

This thesis experimented with survival modeling on ADNI and MVAS and a combination of the two. It showed promising performance on the ADNI dataset with MCI subjects, with a C-index of 0.857, nearing state-of-the-art performance. Modeling on the entire ADNI dataset proved to be inapplicable for Cox-based models, with artificially boosted performances due to the skew in events, even with penalty terms introduced.

The MVAS dataset did not provide the results hoped for, neither as a standalone dataset nor in combination with ADNI. The microvascular biomarkers did not show any better predictive power than other biomarkers, rather quite the opposite. However, this could be due to multiple factors, such as the high amount of missing values and the inherent inability of ADNI3 as a standalone dataset. Thus, the microvascular biomarkers could still be of interest in future research.

7.3 Future Work

Further research should be conducted in the MVAS dataset, perhaps trying other combinations of ADNI cohorts or even completely different datasets. The microvascular biomarkers have shown to be a large part of AD and, in turn, should not be dismissed so easily.

A detriment of the survival models used is they only make predictions based on baseline measurement and do not consider the longitudinal data. This results in removal of all data after baseline visits, which results in a significant loss of information about time varying changes. Using a model that can handle longitudinal data would be very beneficial. One model that can handle longitudinal data recurrently and evaluates based on changes over time is TransformerJM, proposed by [37]. In their original paper, only 11 features were used. Testing the model in a higher dimensional data setting could be of significant interest in the future research.

Bibliography

- [1] A. Gustavsson, N. Norton, T. Fast, *et al.*, “Global estimates on the number of persons across the Alzheimers disease continuum,” *Alzheimer’s Dementia*, vol. 19, pp. 658–670, DOI: [10.1002/alz.12694](https://doi.org/10.1002/alz.12694). [Online]. Available: <https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12694> (visited on 02/19/2024).
- [2] M. Calabrò, C. Rinaldi, G. Santoro, and C. Crisafulli, “The biological pathways of Alzheimer disease: A review,” *AIMS Neuroscience*, vol. 8, no. 1, pp. 86–132, Dec. 16, 2020, ISSN: 2373-8006. DOI: [10.3934/Neuroscience.2021005](https://doi.org/10.3934/Neuroscience.2021005). pmid: [33490374](https://pubmed.ncbi.nlm.nih.gov/33490374/). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7815481/> (visited on 02/19/2024).
- [3] ALZHEIMER’S ASSOCIATION, “2023 Alzheimer’s disease facts and figures,” *Alzheimer’s & Dementia*, vol. 19, no. 4, pp. 1598–1695, 2023, ISSN: 1552-5279. DOI: [10.1002/alz.13016](https://doi.org/10.1002/alz.13016). [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.13016> (visited on 03/03/2024).
- [4] H. Habehh and S. Gohel, “Machine Learning in Healthcare,” *Current Genomics*, vol. 22, no. 4, pp. 291–300, Dec. 16, 2021, ISSN: 1389-2029. DOI: [10.2174/138920292266210705124359](https://doi.org/10.2174/138920292266210705124359). pmid: [35273459](https://pubmed.ncbi.nlm.nih.gov/35273459/). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8822225/> (visited on 05/25/2024).
- [5] M. Tanveer, B. Richhariya, R. U. Khan, *et al.*, “Machine Learning Techniques for the Diagnosis of Alzheimers Disease: A Review,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, 30:1–30:35, 1s Apr. 17, 2020, ISSN: 1551-6857. DOI: [10.1145/3344998](https://doi.org/10.1145/3344998). [Online]. Available: <https://dl.acm.org/doi/10.1145/3344998> (visited on 02/06/2024).
- [6] L. S. Madsen, R. B. Nielsen, P. Parbo, *et al.*, “Capillary function progressively deteriorates in prodromal Alzheimer’s disease: A longitudinal MRI perfusion study,” *Aging Brain*, vol. 2, p. 100035, 2022, ISSN: 2589-9589. DOI: [10.1016/j.nbas.2022.100035](https://doi.org/10.1016/j.nbas.2022.100035). pmid: [36908896](https://pubmed.ncbi.nlm.nih.gov/36908896/).
- [7] M. J. Zigmond, L. P. Rowland, and J. T. Coyle, “Chapter 21: Alzheimer Disease,” in *Neurobiology of Brain Disorders: Biological Basis of Neurological and Psychiatric Disorders*, Amsterdam: Elsevier/AP, 2015, ISBN: 978-0-12-398270-4. [Online]. Available: <https://doi.org/10.1016/B978-0-12-398270-4.00021-5>.

- [8] P. Scheltens, B. De Strooper, M. Kivipelto, *et al.*, “Alzheimer’s disease,” *The Lancet*, vol. 397, no. 10284, pp. 1577–1590, Apr. 2021, ISSN: 01406736. DOI: [10.1016/S0140-6736\(20\)32205-4](https://doi.org/10.1016/S0140-6736(20)32205-4). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0140673620322054> (visited on 02/16/2024).
- [9] G. M. McKhann, D. S. Knopman, H. Chertkow, *et al.*, “The diagnosis of dementia due to Alzheimers disease: Recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimers disease,” *Alzheimer’s & dementia : the journal of the Alzheimer’s Association*, vol. 7, no. 3, pp. 263–269, May 2011, ISSN: 1552-5260. DOI: [10.1016/j.jalz.2011.03.005](https://doi.org/10.1016/j.jalz.2011.03.005). pmid: 21514250. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3312024/> (visited on 05/25/2024).
- [10] “Aducanumab to Be Discontinued as an Alzheimers Treatment,” Alzheimer’s Disease and Dementia. (), [Online]. Available: <https://alz.org/alzheimers-dementia/treatments/aducanumab> (visited on 03/19/2024).
- [11] “Three promising drugs for treating Alzheimer’s disease bring fresh hope | Alzheimer’s Society.” (), [Online]. Available: <https://www.alzheimers.org.uk/blog/three-promising-drugs-for-treating-alzheimers-disease-bring-fresh-hope> (visited on 03/19/2024).
- [12] “Lecanemab Approved for Treatment of Early Alzheimer’s Disease,” Alzheimer’s Disease and Dementia. (), [Online]. Available: <https://alz.org/alzheimers-dementia/treatments/lecanemab-leqembi> (visited on 03/19/2024).
- [13] “Cognitive Health & Digital Biomarker Technologies | Altoida.” (Jan. 22, 2022), [Online]. Available: <https://altoida.com/company/> (visited on 03/19/2024).
- [14] “BrainScope - Team.” (), [Online]. Available: <https://www.brainscope.com/about> (visited on 03/19/2024).
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc., 2012. [Online]. Available: https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html (visited on 05/25/2024).
- [16] R. C. Petersen, P. S. Aisen, L. A. Beckett, *et al.*, “Alzheimer’s Disease Neuroimaging Initiative (ADNI),” *Neurology*, vol. 74, no. 3, pp. 201–209, Jan. 19, 2010, ISSN: 0028-3878. DOI: [10.1212/WNL.0b013e3181cb3e25](https://doi.org/10.1212/WNL.0b013e3181cb3e25). pmid: 20042704. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2809036/> (visited on 06/03/2024).
- [17] D. L. Beekly, E. M. Ramos, W. W. Lee, *et al.*, “The National Alzheimer’s Coordinating Center (NACC) Database: The Uniform Data Set,” *Alzheimer Disease & Associated Disorders*, vol. 21, no. 3, p. 249, Jul.–Sep. 2007, ISSN: 0893-0341. DOI: [10.1097/WAD.0b013e318142774e](https://doi.org/10.1097/WAD.0b013e318142774e). [Online]. Available:

- https://journals.lww.com/alzheimerjournal/abstract/2007/07000/the_national_alzheimer_s_coordinating_center.9.aspx (visited on 05/25/2024).
- [18] L. Liu, S. Zhao, H. Chen, and A. Wang, “A new machine learning method for identifying Alzheimer’s disease,” *Simulation Modelling Practice and Theory*, vol. 99, p. 102023, Feb. 1, 2020, ISSN: 1569-190X. DOI: [10.1016/j.simpat.2019.102023](https://doi.org/10.1016/j.simpat.2019.102023). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1569190X19301546> (visited on 02/07/2024).
- [19] National Center for Biotechnology Information. “PubMed,” PubMed. (), [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/about/> (visited on 03/19/2024).
- [20] A. Spooner, E. Chen, A. Sowmya, *et al.*, “A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction,” *Scientific Reports*, vol. 10, no. 1, p. 20410, 1 Nov. 23, 2020, ISSN: 2045-2322. DOI: [10.1038/s41598-020-77220-w](https://doi.org/10.1038/s41598-020-77220-w). [Online]. Available: <https://www.nature.com/articles/s41598-020-77220-w> (visited on 02/28/2024).
- [21] R. Lee, H. Choi, K.-Y. Park, J.-M. Kim, and J. W. Seok, “Prediction of post-stroke cognitive impairment using brain FDG PET: Deep learning-based approach,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 49, no. 4, pp. 1254–1262, Mar. 1, 2022, ISSN: 1619-7089. DOI: [10.1007/s00259-021-05556-0](https://doi.org/10.1007/s00259-021-05556-0). [Online]. Available: <https://doi.org/10.1007/s00259-021-05556-0> (visited on 05/25/2024).
- [22] Y. Wang, M. Li, L. Kazis, and W. Xia, “The Comparative Effectiveness of Monotherapy and Combination Therapies: Impact of Angiotensin Receptor Blockers on the Onset of Alzheimers Disease,” *JAR life*, vol. 12, pp. 35–45, ISSN: 2534-773X. DOI: [10.14283/jarlife.2023.8](https://doi.org/10.14283/jarlife.2023.8). pmid: 37441415. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10333644/> (visited on 05/25/2024).
- [23] F. Biondo, A. Jewell, M. Pritchard, *et al.*, “Brain-age is associated with progression to dementia in memory clinic patients,” *NeuroImage : Clinical*, vol. 36, p. 103175, Aug. 30, 2022, ISSN: 2213-1582. DOI: [10.1016/j.nicl.2022.103175](https://doi.org/10.1016/j.nicl.2022.103175). pmid: 36087560. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9467894/> (visited on 05/25/2024).
- [24] L. HernándezLorenzo, M. J. GilMoreno, I. OrtegaMadueño, *et al.*, “A datadriven approach to complement the A/T/(N) classification system using CSF biomarkers,” *CNS Neuroscience & Therapeutics*, vol. 30, no. 2, e14382, Jul. 27, 2023, ISSN: 1755-5930. DOI: [10.1111/cns.14382](https://doi.org/10.1111/cns.14382). pmid: 37501389. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10848077/> (visited on 05/25/2024).
- [25] F. Abuhantash, A. A. Shehhi, L. Hadjileontiadis, and M. L. Seghier, “Effect of Comorbidities Features in Machine Learning Models for Survival Analysis to Predict Prodromal Alzheimers Disease,” in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*,

- Jul. 2023, pp. 1–4. DOI: [10.1109/EMBC40787.2023.10341171](https://doi.org/10.1109/EMBC40787.2023.10341171). [Online]. Available: <https://ieeexplore.ieee.org/document/10341171> (visited on 05/19/2024).
- [26] D. Aschwanden, S. Aichele, P. Ghisletta, *et al.*, “Predicting Cognitive Impairment and Dementia: A Machine Learning Approach,” *Journal of Alzheimer’s disease : JAD*, vol. 75, no. 3, pp. 717–728, 2020, ISSN: 1387-2877. DOI: [10.3233/JAD-190967](https://doi.org/10.3233/JAD-190967). pmid: [32333585](#). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7934087/> (visited on 05/20/2024).
- [27] A. Sarica, F. Aracri, M. G. Bianco, M. G. Vaccaro, A. Quattrone, and A. Quattrone, “Conversion from Mild Cognitive Impairment to Alzheimers Disease: A Comparison of Tree-Based Machine Learning Algorithms for Survival Analysis,” in *Brain Informatics*, F. Liu, Y. Zhang, H. Kuai, E. P. Stephen, and H. Wang, Eds., Cham: Springer Nature Switzerland, 2023, pp. 179–190, ISBN: 978-3-031-43075-6. DOI: [10.1007/978-3-031-43075-6_16](https://doi.org/10.1007/978-3-031-43075-6_16).
- [28] A. Sarica, F. Aracri, M. G. Bianco, F. Arcuri, A. Quattrone, and A. Quattrone, “Explainability of random survival forests in predicting conversion risk from mild cognitive impairment to Alzheimers disease,” *Brain Informatics*, vol. 10, no. 1, p. 31, Nov. 18, 2023, ISSN: 2198-4018. DOI: [10.1186/s40708-023-00211-w](https://doi.org/10.1186/s40708-023-00211-w). pmid: [37979033](#). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10657350/> (visited on 05/25/2024).
- [29] S. Song, B. Asken, M. J. Armstrong, Y. Yang, and Z. Li, “Predicting Progression to Clinical Alzheimers Disease Dementia Using the Random Survival Forest,” *Journal of Alzheimer’s disease : JAD*, vol. 95, no. 2, pp. 535–548, 2023, ISSN: 1387-2877. DOI: [10.3233/JAD-230208](https://doi.org/10.3233/JAD-230208). pmid: [37545237](#). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10529100/> (visited on 03/20/2024).
- [30] B. Khajehpiri, H. A. Moghaddam, M. Forouzanfar, *et al.*, “Survival Analysis in Cognitively Normal Subjects and in Patients with Mild Cognitive Impairment Using a Proportional Hazards Model with Extreme Gradient Boosting Regression,” *Journal of Alzheimer’s Disease*, vol. 85, no. 2, pp. 837–850, Jan. 1, 2022, ISSN: 1387-2877. DOI: [10.3233/JAD-215266](https://doi.org/10.3233/JAD-215266). [Online]. Available: <https://content.iospress.com/articles/journal-of-alzheimers-disease/jad215266> (visited on 05/20/2024).
- [31] R. Sharma, H. Anand, Y. Badr, and R. G. Qiu, “Time-to-event prediction using survival analysis methods for Alzheimer’s disease progression,” *Alzheimer’s & Dementia: Translational Research & Clinical Interventions*, vol. 7, no. 1, e12229, 2021, ISSN: 2352-8737. DOI: [10.1002/trc2.12229](https://doi.org/10.1002/trc2.12229). [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/trc2.12229> (visited on 03/24/2024).
- [32] T. Nakagawa, M. Ishida, J. Naito, A. Nagai, S. Yamaguchi, and K. Onoda, “Prediction of conversion to Alzheimers disease using deep survival analysis of MRI images,” *Brain Communications*, vol. 2, no. 1, fcaa057, May 27, 2020, ISSN: 2632-1297. DOI: [10.1093/braincomms/fcaa057](https://doi.org/10.1093/braincomms/fcaa057). pmid: [32954307](#). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7425528/> (visited on 03/24/2024).

- [33] X. Wu, C. Peng, P. T. Nelson, and Q. Cheng, “Machine Learning Approach Predicts Probability of Time to Stage-Specific Conversion of Alzheimers Disease,” *Journal of Alzheimer’s disease : JAD*, vol. 90, no. 2, pp. 891–903, 2022, ISSN: 1387-2877. DOI: [10.3233/JAD-220590](https://doi.org/10.3233/JAD-220590). pmid: 36189595. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9840816/> (visited on 05/19/2024).
- [34] G. Mirabnahrazam, D. Ma, C. Beaulac, *et al.*, “Predicting time-to-conversion for dementia of Alzheimer’s type using multi-modal deep survival analysis,” *Neurobiology of Aging*, vol. 121, pp. 139–156, Jan. 1, 2023, ISSN: 0197-4580. DOI: [10.1016/j.neurobiolaging.2022.10.005](https://doi.org/10.1016/j.neurobiolaging.2022.10.005). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0197458022002196> (visited on 03/18/2024).
- [35] B. Farnsworth von Cederwald, M. Josefsson, A. Wåhlin, L. Nyberg, and N. Karalija, “Association of Cardiovascular Risk Trajectory With Cognitive Decline and Incident Dementia,” *Neurology*, vol. 98, no. 20, e2013–e2022, May 17, 2022. DOI: [10.1212/WNL.00000000000200255](https://doi.org/10.1212/WNL.00000000000200255). [Online]. Available: <https://www.neurology.org/doi/10.1212/WNL.00000000000200255> (visited on 05/20/2024).
- [36] H. Musto, D. Stamate, I. Pu, and D. Stahl, “Predicting Alzheimers Disease Diagnosis Risk Over Time with Survival Machine Learning on the ADNI Cohort,” in *Computational Collective Intelligence*, N. T. Nguyen, J. Botzheim, L. Gulyás, *et al.*, Eds., Cham: Springer Nature Switzerland, 2023, pp. 700–712, ISBN: 978-3-031-41456-5. DOI: [10.1007/978-3-031-41456-5_53](https://doi.org/10.1007/978-3-031-41456-5_53).
- [37] J. Lin and S. Luo, “Deep learning for the dynamic prediction of multivariate longitudinal and survival data,” *Statistics in medicine*, vol. 41, no. 15, pp. 2894–2907, Jul. 10, 2022, ISSN: 0277-6715. DOI: [10.1002/sim.9392](https://doi.org/10.1002/sim.9392). pmid: 35347750. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9232978/> (visited on 05/20/2024).
- [38] Z. Breijyeh and R. Karaman, “Comprehensive Review on Alzheimer’s Disease: Causes and Treatment,” *Molecules (Basel, Switzerland)*, vol. 25, no. 24, p. 5789, Dec. 8, 2020, ISSN: 1420-3049. DOI: [10.3390/molecules25245789](https://doi.org/10.3390/molecules25245789). pmid: 33302541.
- [39] M. A. DeTure and D. W. Dickson, “The neuropathological diagnosis of Alzheimers disease,” *Molecular Neurodegeneration*, vol. 14, no. 1, p. 32, Aug. 2, 2019, ISSN: 1750-1326. DOI: [10.1186/s13024-019-0333-5](https://doi.org/10.1186/s13024-019-0333-5). [Online]. Available: <https://doi.org/10.1186/s13024-019-0333-5> (visited on 02/19/2024).
- [40] H. I. Zeliger, “Chapter 23: Alzheimers disease,” in *Oxidative Stress: Its Mechanisms and Impacts on Human Health and Disease Onset*, London, United Kingdom ; San Diego, CA: Academic Press, an imprint of Elsevier, 2023, ISBN: 978-0-323-91890-9. [Online]. Available: <https://doi.org/10.1016/B978-0-323-91890-9.00020-9>.

- [41] C.-C. Liu, T. Kanekiyo, H. Xu, and G. Bu, “Apolipoprotein E and Alzheimer disease: Risk, mechanisms and therapy,” *Nature Reviews Neurology*, vol. 9, no. 2, pp. 106–118, 2 Feb. 2013, ISSN: 1759-4766. DOI: [10.1038/nrneurol.2012.263](https://doi.org/10.1038/nrneurol.2012.263). [Online]. Available: <https://www.nature.com/articles/nrneurol.2012.263> (visited on 02/17/2024).
- [42] Society for Neuroscience, *Brain Facts: A Primer on the Brain and Nervous System*, 8th ed. 2018. [Online]. Available: <https://www.brainfacts.org/the-brain-facts-book>.
- [43] B. U. Forstmann, M. C. Keuken, and A. Alkemade, “An Introduction to Human Brain Anatomy,” in *An Introduction to Model-Based Cognitive Neuroscience*, B. U. Forstmann and E.-J. Wagenmakers, Eds., New York, NY: Springer, 2015, pp. 71–89, ISBN: 978-1-4939-2236-9. DOI: [10.1007/978-1-4939-2236-9_4](https://doi.org/10.1007/978-1-4939-2236-9_4). [Online]. Available: https://doi.org/10.1007/978-1-4939-2236-9_4 (visited on 05/26/2024).
- [44] D. Gupta, “Chapter 1 - Neuroanatomy,” in *Essentials of Neuroanesthesia*, H. Prabhakar, Ed., Academic Press, Jan. 1, 2017, pp. 3–40, ISBN: 978-0-12-805299-0. DOI: [10.1016/B978-0-12-805299-0.00001-4](https://doi.org/10.1016/B978-0-12-805299-0.00001-4). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128052990000014> (visited on 05/26/2024).
- [45] J. Moini, J. Koenitzer, and A. LoGalbo, “Chapter 1 - Brain structures and functions,” in *Global Emergency of Mental Disorders*, J. Moini, J. Koenitzer, and A. LoGalbo, Eds., Academic Press, Jan. 1, 2021, pp. 3–30, ISBN: 978-0-323-85837-3. DOI: [10.1016/B978-0-323-85837-3.00023-6](https://doi.org/10.1016/B978-0-323-85837-3.00023-6). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780323858373000236> (visited on 05/26/2024).
- [46] “Medical gallery of Blausen Medical 2014,” *WikiJournal of Medicine*, vol. 1, no. 2, 2014, ISSN: 20024436. DOI: [10.15347/wjm/2014.010](https://doi.org/10.15347/wjm/2014.010). [Online]. Available: https://en.wikiversity.org/wiki/WikiJournal_of_Medicine/Medical_gallery_of_Blausen_Medical_2014 (visited on 06/04/2024).
- [47] X. Du, X. Wang, and M. Geng, “Alzheimers disease hypothesis and related therapies,” *Translational Neurodegeneration*, vol. 7, p. 2, Jan. 30, 2018, ISSN: 2047-9158. DOI: [10.1186/s40035-018-0107-y](https://doi.org/10.1186/s40035-018-0107-y). pmid: [29423193](#). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5789526/> (visited on 02/22/2024).
- [48] M. Agarwal, M. R. Alam, M. K. Haider, M. Z. Malik, and D.-K. Kim, “Alzheimers Disease: An Overview of Major Hypotheses and Therapeutic Options in Nanotechnology,” *Nanomaterials*, vol. 11, no. 1, p. 59, Dec. 29, 2020, ISSN: 2079-4991. DOI: [10.3390/nano11010059](https://doi.org/10.3390/nano11010059). pmid: [33383712](#). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7823376/> (visited on 02/22/2024).

- [49] L. Pini, M. Pievani, M. Bocchetta, *et al.*, “Brain atrophy in Alzheimers Disease and aging,” *Ageing Research Reviews*, Brain Imaging and Aging, vol. 30, pp. 25–48, Sep. 1, 2016, ISSN: 1568-1637. DOI: [10.1016/j.arr.2016.01.002](https://doi.org/10.1016/j.arr.2016.01.002). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1568163716300022> (visited on 05/26/2024).
- [50] G.-f. Chen, T.-h. Xu, Y. Yan, *et al.*, “Amyloid beta: Structure, biology and structure-based therapeutic development,” *Acta Pharmacologica Sinica*, vol. 38, no. 9, pp. 1205–1235, 9 Sep. 2017, ISSN: 1745-7254. DOI: [10.1038/aps.2017.28](https://doi.org/10.1038/aps.2017.28). [Online]. Available: <https://www.nature.com/articles/aps201728> (visited on 02/18/2024).
- [51] N. Korte, R. Nortley, and D. Attwell, “Cerebral blood flow decrease as an early pathological mechanism in Alzheimer’s disease,” *Acta Neuropathologica*, vol. 140, no. 6, pp. 793–810, Dec. 1, 2020, ISSN: 1432-0533. DOI: [10.1007/s00401-020-02215-w](https://doi.org/10.1007/s00401-020-02215-w). [Online]. Available: <https://doi.org/10.1007/s00401-020-02215-w> (visited on 02/18/2024).
- [52] “What Happens to the Brain in Alzheimer’s Disease?” National Institute on Aging. (), [Online]. Available: <https://www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-happens-brain-alzheimers-disease> (visited on 02/16/2024).
- [53] A. Lloret, M.-C. Badia, E. Giraldo, *et al.*, “Amyloid- toxicity and tau hyperphosphorylation are linked via RCAN1 in Alzheimer’s disease,” *Journal of Alzheimer’s disease: JAD*, vol. 27, no. 4, pp. 701–709, 2011, ISSN: 1875-8908. DOI: [10.3233/JAD-2011-110890](https://doi.org/10.3233/JAD-2011-110890). pmid: [21876249](https://pubmed.ncbi.nlm.nih.gov/21876249/).
- [54] K. K. F. Tsoi, J. Y. C. Chan, H. W. Hirai, S. Y. S. Wong, and T. C. Y. Kwok, “Cognitive Tests to Detect Dementia: A Systematic Review and Meta-analysis,” *JAMA Internal Medicine*, vol. 175, no. 9, pp. 1450–1458, Sep. 1, 2015, ISSN: 2168-6106. DOI: [10.1001/jamainternmed.2015.2152](https://doi.org/10.1001/jamainternmed.2015.2152). [Online]. Available: <https://doi.org/10.1001/jamainternmed.2015.2152> (visited on 05/27/2024).
- [55] I. ArevaloRodriguez, N. Smailagic, M. R. i Figuls, *et al.*, “MiniMental State Examination (MMSE) for the detection of Alzheimer’s disease and other dementias in people with mild cognitive impairment (MCI),” *Cochrane Database of Systematic Reviews*, no. 3, 2015, ISSN: 1465-1858. DOI: [10.1002/14651858.CD010783.pub2](https://doi.org/10.1002/14651858.CD010783.pub2). [Online]. Available: <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD010783.pub2/full> (visited on 05/25/2024).
- [56] Andrew J. Larner, *Cognitive Screening Instruments*, 2nd ed. New York, NY: Springer Berlin Heidelberg, 2016, ISBN: 978-3-319-44774-2. [Online]. Available: <https://doi.org/10.1007/978-3-319-44775-9>.
- [57] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, *et al.*, “The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment,” *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, 2005, ISSN: 1532-5415. DOI: [10.1111/j.1532-5415.2005.53221.x](https://doi.org/10.1111/j.1532-5415.2005.53221.x). [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1532-5415.2005.53221.x> (visited on 05/27/2024).

- [58] M. M. Williams, M. Storandt, C. M. Roe, and J. C. Morris, “Progression of Alzheimers disease as measured by Clinical Dementia Rating Sum of Boxes scores,” *Alzheimer’s & Dementia*, vol. 9, S39–S44, 1, Supplement Feb. 1, 2013, ISSN: 1552-5260. doi: [10.1016/j.jalz.2012.01.005](https://doi.org/10.1016/j.jalz.2012.01.005). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1552526012000246> (visited on 05/27/2024).
- [59] J. K. Kueper, M. Speechley, and M. Montero-Odasso, “The Alzheimers Disease Assessment ScaleCognitive Subscale (ADAS-Cog): Modifications and Responsiveness in Pre-Dementia Populations. A Narrative Review,” *Journal of Alzheimer’s Disease*, vol. 63, no. 2, pp. 423–444, ISSN: 1387-2877. doi: [10.3233/JAD-170991](https://doi.org/10.3233/JAD-170991). pmid: [29660938](https://pubmed.ncbi.nlm.nih.gov/29660938/). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5929311/> (visited on 05/25/2024).
- [60] “Report Viewer | NINDS Common Data Elements.” (), [Online]. Available: [https://www.commondataelements.ninds.nih.gov/report-viewer/23949/Rey%20Auditory%20Verbal%20Learning%20Test%20\(RAVLT\)](https://www.commondataelements.ninds.nih.gov/report-viewer/23949/Rey%20Auditory%20Verbal%20Learning%20Test%20(RAVLT)) (visited on 05/27/2024).
- [61] K. Ito, M. M. Hutmacher, and B. W. Corrigan, “Modeling of Functional Assessment Questionnaire (FAQ) as continuous bounded data from the ADNI database,” *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 39, no. 6, pp. 601–618, Dec. 1, 2012, ISSN: 1573-8744. doi: [10.1007/s10928-012-9271-3](https://doi.org/10.1007/s10928-012-9271-3). [Online]. Available: <https://doi.org/10.1007/s10928-012-9271-3> (visited on 05/27/2024).
- [62] S. T. Farias, D. Mungas, B. R. Reed, *et al.*, “The measurement of everyday cognition (ECog): Scale development and psychometric properties,” *Neuropsychology*, vol. 22, no. 4, pp. 531–544, 2008, ISSN: 1931-1559. doi: [10.1037/0894-4105.22.4.531](https://doi.org/10.1037/0894-4105.22.4.531).
- [63] K. R. Chapman, H. Bing-Canar, M. L. Alosco, *et al.*, “Mini Mental State Examination and Logical Memory scores for entry into Alzheimers disease trials,” *Alzheimer’s Research & Therapy*, vol. 8, p. 9, Feb. 22, 2016, ISSN: 1758-9193. doi: [10.1186/s13195-016-0176-z](https://doi.org/10.1186/s13195-016-0176-z). pmid: [26899835](https://pubmed.ncbi.nlm.nih.gov/26899835/). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4762168/> (visited on 05/27/2024).
- [64] J. Jaeger, “Digit Symbol Substitution Test,” *Journal of Clinical Psychopharmacology*, vol. 38, no. 5, pp. 513–519, Oct. 2018, ISSN: 0271-0749. doi: [10.1097/JCP.0000000000000941](https://doi.org/10.1097/JCP.0000000000000941). pmid: [30124583](https://pubmed.ncbi.nlm.nih.gov/30124583/). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6291255/> (visited on 05/27/2024).
- [65] “Cerebral Blood Flow - an overview | ScienceDirect Topics.” (), [Online]. Available: <https://www.sciencedirect.com/topics/psychology/cerebral-blood-flow> (visited on 02/28/2024).

- [66] F. Gaillard. “Cerebral blood volume (CBV) | Radiology Reference Article | Radiopaedia.org,” Radiopaedia. (), [Online]. Available: <https://radiopaedia.org/articles/cerebral-blood-volume-cbv> (visited on 02/28/2024).
- [67] B. D. Muzio. “Mean transit time (MTT) | Radiology Reference Article | Radiopaedia.org,” Radiopaedia. (), [Online]. Available: <https://radiopaedia.org/articles/mean-transit-time-mtt?lang=us> (visited on 02/28/2024).
- [68] A. K. Dotiwala, C. McCausland, and N. S. Samra, “Anatomy, Head and Neck: Blood Brain Barrier,” in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2024. pmid: 30137840. [Online]. Available: <http://www.ncbi.nlm.nih.gov/books/NBK519556/> (visited on 02/28/2024).
- [69] V. De Santis and M. Singer, “Tissue oxygen tension monitoring of organ perfusion: Rationale, methodologies, and literature review,” *British Journal of Anaesthesia*, vol. 115, no. 3, pp. 357–365, Sep. 1, 2015, ISSN: 0007-0912. DOI: [10.1093/bja/aev162](https://doi.org/10.1093/bja/aev162). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0007091217311480> (visited on 02/29/2024).
- [70] J. D. Sengillo, E. A. Winkler, C. T. Walker, J. S. Sullivan, M. Johnson, and B. V. Zlokovic, “Deficiency in Mural Vascular Cells Coincides with BloodBrain Barrier Disruption in Alzheimer’s Disease,” *Brain Pathology*, vol. 23, no. 3, pp. 303–310, 2013, ISSN: 1750-3639. DOI: [10.1111/bpa.12004](https://doi.org/10.1111/bpa.12004). [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/bpa.12004> (visited on 02/29/2024).
- [71] B. V. Zlokovic, “Neurovascular pathways to neurodegeneration in Alzheimer’s disease and other disorders,” *Nature Reviews Neuroscience*, vol. 12, no. 12, pp. 723–738, 12 Dec. 2011, ISSN: 1471-0048. DOI: [10.1038/nrn3114](https://doi.org/10.1038/nrn3114). [Online]. Available: <https://www.nature.com/articles/nrn3114> (visited on 02/29/2024).
- [72] L. Østergaard, R. Aamand, E. Gutiérrez-Jiménez, *et al.*, “The capillary dysfunction hypothesis of Alzheimer’s disease,” *Neurobiology of Aging*, vol. 34, no. 4, pp. 1018–1031, Apr. 1, 2013, ISSN: 0197-4580. DOI: [10.1016/j.neurobiolaging.2012.09.011](https://doi.org/10.1016/j.neurobiolaging.2012.09.011). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0197458012004678> (visited on 02/20/2024).
- [73] L. S. Madsen, P. Parbo, R. Ismail, *et al.*, “Capillary dysfunction correlates with cortical amyloid load in early Alzheimer’s disease,” *Neurobiology of Aging*, vol. 123, pp. 1–9, Mar. 1, 2023, ISSN: 0197-4580. DOI: [10.1016/j.neurobiolaging.2022.12.006](https://doi.org/10.1016/j.neurobiolaging.2022.12.006). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0197458022002603> (visited on 01/14/2024).
- [74] O. Islam, “Alzheimer Disease Imaging,” 2022-12-28-12:57. [Online]. Available: <https://emedicine.medscape.com/article/336281-overview?form=fpf> (visited on 03/05/2024).

- [75] J. Kim, M. Jeong, W. R. Stiles, and H. S. Choi, “Neuroimaging Modalities in Alzheimers Disease: Diagnosis and Clinical Features,” *International Journal of Molecular Sciences*, vol. 23, no. 11, p. 6079, 11 Jan. 2022, ISSN: 1422-0067. DOI: [10.3390/ijms23116079](https://doi.org/10.3390/ijms23116079). [Online]. Available: <https://www.mdpi.com/1422-0067/23/11/6079> (visited on 03/05/2024).
- [76] F. Gaillard. “MRI sequences (overview) | Radiology Reference Article | Radiopaedia.org,” Radiopaedia. (), [Online]. Available: <https://radiopaedia.org/articles/mri-sequences-overview> (visited on 02/14/2024).
- [77] N. D. DiProspero, S. Kim, and M. A. Yassa, “Chapter 8 - Magnetic resonance imaging biomarkers for cognitive decline in Down syndrome,” in *The Neurobiology of Aging and Alzheimer Disease in Down Syndrome*, E. Head and I. Lott, Eds., Academic Press, Jan. 1, 2022, pp. 149–172, ISBN: 978-0-12-818845-3. DOI: [10.1016/B978-0-12-818845-3.00014-1](https://doi.org/10.1016/B978-0-12-818845-3.00014-1). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128188453000141> (visited on 03/06/2024).
- [78] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, “Brain Imaging in Alzheimer Disease,” *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 4, a006213, Apr. 2012, ISSN: 2157-1422. DOI: [10.1101/cshperspect.a006213](https://doi.org/10.1101/cshperspect.a006213). pmid: [22474610](https://pubmed.ncbi.nlm.nih.gov/22474610/). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3312396/> (visited on 02/14/2024).
- [79] K. Welker, J. Boxerman, A. Kalnin, *et al.*, “ASFNR Recommendations for Clinical Performance of MR Dynamic Susceptibility Contrast Perfusion Imaging of the Brain,” *American Journal of Neuroradiology*, vol. 36, no. 6, E41–E51, Jun. 1, 2015, ISSN: 0195-6108, 1936-959X. DOI: [10.3174/ajnr.A4341](https://doi.org/10.3174/ajnr.A4341). pmid: [25907520](https://pubmed.ncbi.nlm.nih.gov/25907520/). [Online]. Available: <https://www.ajnr.org/content/36/6/E41> (visited on 03/08/2024).
- [80] J. D. Kropotov, “Chapter 1.4 - Positron Emission Tomography,” in *Functional Neuromarkers for Psychiatry*, J. D. Kropotov, Ed., San Diego: Academic Press, Jan. 1, 2016, pp. 27–30, ISBN: 978-0-12-410513-3. DOI: [10.1016/B978-0-12-410513-3.00004-8](https://doi.org/10.1016/B978-0-12-410513-3.00004-8). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780124105133000048> (visited on 03/13/2024).
- [81] E. T. Lee and J. W. Wang, *Statistical Methods for Survival Data Analysis* (Wiley Series in Probability and Statistics), 1st ed. Wiley, Apr. 4, 2003, ISBN: 978-0-471-36997-4 978-0-471-45854-8. DOI: [10.1002/0471458546](https://doi.org/10.1002/0471458546). [Online]. Available: <https://onlinelibrary.wiley.com/doi/book/10.1002/0471458546> (visited on 05/13/2024).
- [82] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, *An Introduction to Statistical Learning*. Cham: Springer International Publishing, 2023, ISBN: 978-3-031-38747-0. DOI: [10.1007/978-3-031-38747-0_1](https://doi.org/10.1007/978-3-031-38747-0_1). [Online]. Available: <https://www.statlearning.com/> (visited on 05/13/2024).

- [83] D. R. Cox, “Regression Models and Life-Tables,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, Jan. 1, 1972, ISSN: 0035-9246. DOI: [10.1111/j.2517-6161.1972.tb00899.x](https://doi.org/10.1111/j.2517-6161.1972.tb00899.x). [Online]. Available: <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x> (visited on 05/12/2024).
- [84] G. Ambler, S. Seaman, and R. Z. Omar, “An evaluation of penalised survival methods for developing prognostic models with rare events,” *Statistics in Medicine*, vol. 31, no. 11-12, pp. 1150–1161, 2012, ISSN: 1097-0258. DOI: [10.1002/sim.4371](https://doi.org/10.1002/sim.4371). [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4371> (visited on 06/02/2024).
- [85] H. Zou and T. Hastie, “Regularization and Variable Selection Via the Elastic Net,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, Apr. 1, 2005, ISSN: 1369-7412. DOI: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x). [Online]. Available: <https://doi.org/10.1111/j.1467-9868.2005.00503.x> (visited on 06/02/2024).
- [86] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, “Regularization Paths for Coxs Proportional Hazards Model via Coordinate Descent,” *Journal of statistical software*, vol. 39, no. 5, pp. 1–13, Mar. 2011, ISSN: 1548-7660. DOI: [10.18637/jss.v039.i05](https://doi.org/10.18637/jss.v039.i05). pmid: 27065756. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4824408/> (visited on 06/02/2024).
- [87] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 1, 2001, ISSN: 1573-0565. DOI: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324). [Online]. Available: <https://doi.org/10.1023/A:1010933404324> (visited on 06/01/2024).
- [88] F. E. Harrell, K. L. Lee, and D. B. Mark, “Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors,” *Statistics in Medicine*, vol. 15, no. 4, pp. 361–387, Feb. 28, 1996, ISSN: 0277-6715. DOI: [10.1002/\(SICI\)1097-0258\(19960229\)15:4<361::AID-SIM168>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4). pmid: 8668867.
- [89] H. Uno, T. Cai, M. J. Pencina, R. B. D’Agostino, and L. J. Wei, “On the C-statistics for Evaluating Overall Adequacy of Risk Prediction Procedures with Censored Survival Data,” *Statistics in Medicine*, vol. 30, no. 10, pp. 1105–1117, 2011, ISSN: 1097-0258. DOI: [10.1002/sim.4154](https://doi.org/10.1002/sim.4154). [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4154> (visited on 05/31/2024).
- [90] G. W. Brier, “Verification of forecasts expressed in terms of probability,” *Monthly Weather Review*, vol. 78, no. 1, pp. 1–3, Jan. 1, 1950, ISSN: 1520-0493, 0027-0644. DOI: [10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2). [Online]. Available: https://journals.ametsoc.org/view/journals/mwre/78/1/1520-0493_1950_078_0001_vofeit_2_0_co_2.xml (visited on 05/29/2024).

- [91] E. W. Steyerberg, A. J. Vickers, N. R. Cook, *et al.*, “Assessing the performance of prediction models: A framework for some traditional and novel measures,” *Epidemiology (Cambridge, Mass.)*, vol. 21, no. 1, pp. 128–138, Jan. 2010, ISSN: 1044-3983. DOI: [10.1097/EDE.0b013e3181c30fb2](https://doi.org/10.1097/EDE.0b013e3181c30fb2). pmid: [20010215](#). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3575184/> (visited on 05/31/2024).
- [92] S. Y. Park, J. E. Park, H. Kim, and S. H. Park, “Review of Statistical Methods for Evaluating the Performance of Survival or Other Time-to-Event Prediction Models (from Conventional to Deep Learning Approaches),” *Korean Journal of Radiology*, vol. 22, no. 10, pp. 1697–1707, Oct. 2021, ISSN: 1229-6929. DOI: [10.3348/kjr.2021.0223](https://doi.org/10.3348/kjr.2021.0223). pmid: [34269532](#). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8484151/> (visited on 05/31/2024).
- [93] F. E. Harrell, “Chapter 3: Missing Data,” in *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, ser. Springer Series in Statistics, 2nd edition, Cham: Springer, Mar. 3, 2024, ISBN: 978-3-319-19425-7. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-319-19425-7> (visited on 04/15/2024).
- [94] P. C. Austin, I. R. White, D. S. Lee, and S. van Buuren, “Missing Data in Clinical Research: A Tutorial on Multiple Imputation,” *Canadian Journal of Cardiology*, vol. 37, no. 9, pp. 1322–1331, Sep. 1, 2021, ISSN: 0828-282X. DOI: [10.1016/j.cjca.2020.11.010](https://doi.org/10.1016/j.cjca.2020.11.010). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0828282X20311119> (visited on 04/14/2024).
- [95] D. J. Stekhoven and P. Bühlmann, “MissForestnon-parametric missing value imputation for mixed-type data,” *Bioinformatics*, vol. 28, no. 1, pp. 112–118, Jan. 1, 2012, ISSN: 1367-4803. DOI: [10.1093/bioinformatics/btr597](https://doi.org/10.1093/bioinformatics/btr597). [Online]. Available: <https://doi.org/10.1093/bioinformatics/btr597> (visited on 04/14/2024).
- [96] S. van Buuren, *Flexible Imputation of Missing Data* (Chapman & Hall/CRC Interdisciplinary Statistics Series), Second edition. Boca Raton London New York: CRC Press, Taylor & Francis Group, 2018, 415 pp., ISBN: 978-1-38-58831-8. [Online]. Available: <https://stefvanbuuren.name/fimd/>.
- [97] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, “Multiple imputation by chained equations: What is it and how does it work?” *International Journal of Methods in Psychiatric Research*, vol. 20, no. 1, pp. 40–49, Feb. 24, 2011, ISSN: 1049-8931. DOI: [10.1002/mpr.329](https://doi.org/10.1002/mpr.329). pmid: [21499542](#). [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074241/> (visited on 05/28/2024).
- [98] Y. Luo, “Evaluating the state of the art in missing data imputation for clinical data,” *Briefings in Bioinformatics*, vol. 23, no. 1, bbab489, Jan. 1, 2022, ISSN: 1477-4054. DOI: [10.1093/bib/bbab489](https://doi.org/10.1093/bib/bbab489). [Online]. Available: <https://doi.org/10.1093/bib/bbab489> (visited on 04/14/2024).

- [99] K. Seu, M.-S. Kang, and H. Lee, “An Intelligent Missing Data Imputation Techniques: A Review,” *JOIV : International Journal on Informatics Visualization*, vol. 6, no. 1-2, pp. 278–283, May 31, 2022, ISSN: 2549-9904. DOI: [10.30630/joiv.6.1-2.935](https://doi.org/10.30630/joiv.6.1-2.935). [Online]. Available: <http://joiv.org/index.php/joiv/article/view/935> (visited on 04/14/2024).
- [100] A. Marshall, D. G. Altman, R. L. Holder, and P. Royston, “Combining estimates of interest in prognostic modelling studies after multiple imputation: Current practice and guidelines,” *BMC Medical Research Methodology*, vol. 9, no. 1, p. 57, Jul. 28, 2009, ISSN: 1471-2288. DOI: [10.1186/1471-2288-9-57](https://doi.org/10.1186/1471-2288-9-57). [Online]. Available: <https://doi.org/10.1186/1471-2288-9-57> (visited on 06/03/2024).
- [101] C. R. Harris, K. J. Millman, S. J. van der Walt, *et al.*, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020, ISSN: 1476-4687. DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2). [Online]. Available: <https://www.nature.com/articles/s41586-020-2649-2> (visited on 05/10/2024).
- [102] The pandas development team, *Pandas-dev/pandas: Pandas*, version 2.1.3, 2020. DOI: [10.5281/ZENODO.3509134](https://doi.org/10.5281/ZENODO.3509134). [Online]. Available: <https://zenodo.org/doi/10.5281/zenodo.3509134> (visited on 05/10/2024).
- [103] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011, ISSN: 1533-7928. [Online]. Available: <http://jmlr.org/papers/v12/pedregosa11a.html> (visited on 05/10/2024).
- [104] S. Pölsterl, “Scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn,” *Journal of Machine Learning Research*, vol. 21, no. 212, pp. 1–6, 2020, ISSN: 1533-7928. [Online]. Available: <http://jmlr.org/papers/v21/20-729.html> (visited on 05/10/2024).
- [105] C. Davidson-Pilon, *Lifelines, survival analysis in Python*, version v0.28.0, Zenodo, Jan. 3, 2024. DOI: [10.5281/ZENODO.805993](https://doi.org/10.5281/ZENODO.805993). [Online]. Available: <https://zenodo.org/doi/10.5281/zenodo.805993> (visited on 06/01/2024).
- [106] “ABOUT ADNI.” (), [Online]. Available: <https://adni.loni.usc.edu/about/> (visited on 04/23/2024).
- [107] R. Y. Lo, W. J. Jagust, and Alzheimer’s Disease Neuroimaging Initiative, “Predicting missing biomarker data in a longitudinal study of Alzheimer disease,” *Neurology*, vol. 78, no. 18, pp. 1376–1382, May 1, 2012, ISSN: 1526-632X. DOI: [10.1212/WNL.0b013e318253d5b3](https://doi.org/10.1212/WNL.0b013e318253d5b3). pmid: [22491869](#).
- [108] Alzheimer’s Disease Neuroimaging Initiative. “ADNI | Study Design.” (), [Online]. Available: <https://adni.loni.usc.edu/study-design/> (visited on 06/03/2024).
- [109] S. Wilson, *Miceforest: Missing Value Imputation using LightGBM*, version 5.7.0. [Online]. Available: <https://github.com/AnotherSamWilson/miceforest> (visited on 05/11/2024).

- [110] G. Ke, Q. Meng, T. Finley, *et al.*, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” *Advances in neural information processing systems*, vol. 30, pp. 3146–3154, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (visited on 06/01/2024).
- [111] G. Tondo, L. Iaccarino, C. Cerami, *et al.*, “¹¹C-PK11195 PET-based molecular study of microglia activation in SOD1 amyotrophic lateral sclerosis,” *Annals of Clinical and Translational Neurology*, vol. 7, no. 9, pp. 1513–1523, Sep. 2020, ISSN: 2328-9503. DOI: [10.1002/acn3.51112](https://doi.org/10.1002/acn3.51112). pmid: [32762033](https://pubmed.ncbi.nlm.nih.gov/32762033/).

APPENDIX A

ADNI Dataset

Table A.1: ADNI Merge Dataset Columns Description

Variable name	Datatype	Data missing	Description
RID	Integer	0 (0%)	Participant roster ID
COLPROT	String	0 (0%)	Study protocol of current measurement
ORIGPROT	String (category)	0 (0%)	Baseline study protocol
PTID	String	0 (0%)	Participant ID
SITE	Integer	0 (0%)	Site of data collection
VISCODE	String (category)	0 (0%)	Visit code
EXAMDATE	String (date)	0 (0%)	Examination date
DX_bl	String (category)	11 (0.1%)	Baseline diagnosis
AGE	Continous	9 (0.1%)	Age at baseline
PTGENDER	String (category)	0 (0%)	Gender of participant
PTEDUCAT	Integer	0 (0%)	Years of education
PTETHCAT	String (category)	0 (0%)	If ethnicity is Hispanic/Latino
PTRACCAT	String (category)	0 (0%)	Racial category
PTMARRY	String (category)	1 (0.0%)	Marital status
APOE4	Integer	365 (2.2%)	Number of APOEe4 alleles
FDG	Continous	12728 (77.5%)	values extracted from FDG PET scans for meta regions of interest.
PIB	Continous	16198 (98.6%)	Average PIB SUVR of frontal cortex, anterior cingulate, precuneus cortex, and parietal cortex
AV45	Continous	13341 (81.2%)	Summary florbetapir (binds to A β) cortical SUVR normalized by whole cerebellum.
FBB	Continous	15908 (96.9%)	Summary florbetaben cortical SUVR normalized by whole cerebellum
ABETA	Continous	14051 (85.6%)	$\text{A}\beta$ value from CSF
TAU	Continous	14051 (85.6%)	Tau value from CSF
PTAU	Continous	14052 (85.6%)	PTAU value from CSF
CDRSB	Continous	4674 (28.5%)	Clinical Dementia Rating-Sum of Boxes score
ADAS11	Continous	4990 (30.4%)	Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog)-11 score

Variable name	Datatype	Data missing	Description
ADAS13	Integer/ Continous	5091 (31.0%)	Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog)-13 score, including Delayed Word Recall and Number Cancellation
ADASQ4	Integer	4954 (30.2%)	ADAS Delayed Word Recall score
MMSE	Integer	4952 (30.2%)	Mini-Mental State Examination score
RAVLT_immediate	Integer	5075 (30.9%)	Rey's Auditory Verbal Learning Test Immediate recall score (sum of 5 trials)
RAVLT_learning	Integer	5074 (30.9%)	Rey's Auditory Verbal Learning Test learning score (trial 5 - trial 1)
RAVLT_forgetting	Integer	5106 (31.1%)	Rey's Auditory Verbal Learning Test learning forgetting score (trial 5 - delayed)
RAVLT_perc.forgetting	Integer	5179 (31.5%)	Rey's Auditory Verbal Learning Test learning percent forgetting score
LDELTOTAL	Integer	6979 (42.5%)	Logical memory delaid recall
DIGITSCOR	Integer	12621 (76.9%)	Digit Symbol Substitution Test score
TRABSCOR	Continous/ Integer	5414 (33.0%)	The trail making test, part B
FAQ	Integer	4678 (28.5%)	Functional Assessment Questionnaire score
MOCA	Integer	8995 (54.8%)	Montreal cognitive assessment score
EcogPtMem	Continous	8645 (52.6%)	Everyday Cognition Scale by participant - Memory
EcogPtLang	Continous	8662 (52.7%)	Everyday Cognition Scale by participant - Language
EcogPtVisspat	Continous	8745 (53.3%)	Everyday Cognition Scale by participant - Visual/Spatial
EcogPtPlan	Continous	8681 (52.9%)	Everyday Cognition Scale by participant - Planning
EcogPtOrgan	Continous	8823 (53.7%)	Everyday Cognition Scale by participant - Organization
EcogPtDivatt	Continous	8721 (53.1%)	Everyday Cognition Scale by participant - Divided Attention
EcogPtTotal	Continous	8655 (52.7%)	Everyday Cognition Scale by participant - Total
EcogSPMem	Continous	8602 (52.4%)	Everyday Cognition Scale study partner - Memory
EcogSPLang	Continous	8596 (52.3%)	Everyday Cognition Scale study partner - Language
EcogSPVisspat	Continous	8782 (53.5%)	Everyday Cognition Scale study partner - Visuospatial Abilities
EcogSPPlan	Continous	8686 (52.9%)	Everyday Cognition Scale study partner - Planning
EcogSPOrgan	Continous	8933 (54.4%)	Everyday Cognition Scale study partner - Organization
EcogSPDivatt	Continous	8804 (53.6%)	Everyday Cognition Scale by study partner - Divided Attention
EcogSPTotal	Continous	8604 (52.4%)	Everyday Cognition Scale by study partner - Total
FLDSTRENG	String (category)	8762 (53.4%)	MRI Field Strength
FSVERSION	String (category)	6289 (38.3%)	FreeSurfer Software Version
IMAGEUID	Continous	6289 (38.3%)	Image ID on the LONI website
Ventricles	Continous	6953 (42.3%)	sMRI volume of ventricles

Variable name	Datatype	Data missing	Description
Hippocampus	Continous	7568 (46.1%)	sMRI volume of hippocampus
WholeBrain	Continous	6679 (40.7%)	sMRI volume of the whole brain
Entorhinal	Continous	7981 (48.6%)	sMRI volume of entorhinal brain region
Fusiform	Continous	7981 (48.6%)	sMRI volume of the fusiform gyrus
MidTemp	Continous	7981 (48.6%)	sMRI volume of the middle temporal gyrus
ICV	Continous	6306 (38.4%)	sMRI Intracranial volume
DX	String (category)	4963 (30.2%)	Diagnosis given at visit
mPACCdigit	Continous	4949 (30.1%)	
mPACCtrailsB	Continous	4943 (30.1%)	
EXAMDATE_bl	String (date)	0 (0.0%)	Baseline exam date
CDRSB_bl	Integer	0 (0.0%)	Baseline measurement of CDRSB
ADAS11_bl	Continous	40 (0.2%)	Baseline measurement of ADAS11
ADAS13_bl	Continous	112 (0.7%)	Baseline measurement of ADAS13
ADASQ4_bl	Integer	7 (0.0%)	Baseline measurement of ADASQ4
MMSE_bl	Integer	2 (0.0%)	Baseline measurement of MMSE
RAVLT_immediate_bl	Integer	41 (0.2%)	Baseline measurement of RAVLT_immediate
RAVLT_learning_bl	Integer	41 (0.2%)	Baseline measurement of RAVLT_learning
RAVLT_forgetting_bl	Integer	44 (0.3%)	Baseline measurement of RAVLT_forgetting
RAVLT_perc_forgetting_bl	Continous	65 (0.4%)	Baseline measurement of RAVLT_perc_forgetting
LDELTOTAL_BL	Integer	5 (0.0%)	Baseline measurement of LDELTOTAL
DIGITSCOR_bl	Integer	9171 (55.8%)	Baseline measurement of DIGITSCOR
TRABSCOR_bl	Integer	267 (1.6%)	Baseline measurement of TRABSCOR
FAQ_bl	Integer	86 (0.5%)	Baseline measurement of FAQ
mPACCdigit_bl	Continous	3 (0.0%)	Baseline measurement of mPACCdigit
mPACCtrailsB_bl	Continous	3 (0.0%)	Baseline measurement of mPACCtrailsB
FLDSTRENG_bl	String (category)	3425 (20.9%)	Baseline value of FLDSTRENG
FSVERSION_bl	String (category)	261 (1.6%)	Baseline FSVERSION
IMAGEUID_bl	Continous	261 (1.6%)	baseline IMAGEUID
Ventricles_bl	Continous	778 (4.7%)	Baseline measurement of Ventricles
Hippocampus_bl	Continous	2149 (13.1%)	Baseline measurement of Hippocampus
WholeBrain_bl	Continous	476 (2.9%)	Baseline measurement of WholeBrain
Entorhinal_bl	Continous	2310 (14.1%)	Baseline measurement of Entorhinal
Fusiform_bl	Continous	2310 (14.1%)	Baseline measurement of Fusiform
MidTemp_bl	Continous	2310 (14.1%)	Baseline measurement of MidTemp
ICV_bl	Continous	263 (1.6%)	Baseline measurement of ICV
MOCA_bl	Integer	7423 (45.2%)	Baseline measurement of MOCA
EcogPtMem_bl	Continous	7306 (44.5%)	Baseline measurement of EcogPtMem
EcogPtLang_bl	Continous	7315 (44.5%)	Baseline measurement of EcogPtLang
EcogPtVisspat_bl	Continous	7349 (44.8%)	Baseline measurement of EcogPtVisspat
EcogPtPlan_bl	Continous	7305 (44.5%)	Baseline measurement of EcogPtPlan
EcogPtOrgan_bl	Continous	7489 (45.6%)	Baseline measurement of EcogPtOrgan
EcogPtDivatt_bl	Continous	7361 (44.8%)	Baseline measurement of EcogPtDivatt
EcogPtTotal_bl	Continous	7304 (44.5%)	Baseline measurement of EcogPtTotal
EcogSPMem_bl	Continous	7338 (44.7%)	Baseline measurement of EcogSPMem

Variable name	Datatype	Data missing	Description
EcogSPLang_bl	Continous	7327 (44.6%)	Baseline measurement of EcogSPLang
EcogSPVisspat_bl	Continous	7460 (45.4%)	Baseline measurement of EcogSPCVisspat
EcogSPPlan_bl	Continous	7389 (45.0%)	Baseline measurement of EcogSPPan
EcogSPOrgan_bl	Continous	7742 (47.1%)	Baseline measurement of EcogSPOrgan
EcogSPDivatt_bl	Continous	7506 (45.7%)	Baseline measurement of EcogSPDivatt
EcogSPTotal_bl	Continous	7336 (44.7%)	Baseline measurement of EcogSPTotal
ABETA_bl	Continous	6338 (38.6%)	Baseline measurement of ABETA
TAU_bl	Continous	6338 (38.6%)	Baseline measurement of TAU
PTAU_bl	Continous	6338 (38.6%)	Baseline measurement of PTAU
FDG_bl	Continous	5201 (31.7%)	Baseline measurement of FDG
PIB_bl	Continous	16267 (99.1%)	Baseline measurement of PIB
AV45_bl	Continous	8585 (52.3%)	Baseline measurement of AV45
FBB_bl	Continous	15322 (93.3%)	Baseline measurement of FBB
Years_bl	Continous	0 (0.0%)	Years since baseline measurement
Month_bl	Continous	0 (0.0%)	Months since baseline
Month	Integer	0 (0.0%)	Rounded months since baseline
M	Integer	0 (0.0%)	Months based on VISCODE

Table A.2: ADNI Merge Dataset Columns Description

Variable name	Missing in cohort
PIB	ADNI2, ADNIGO, ADNI3
AV45	ADNI1
FBB	ADNI1, ADNIGO, ADNI2
ABETA	ADNI3
TAU	ADNI3
PTAU	ADNI3
DIGITSCOR	ADNI2, ADNIGO, ADNI3
MOCA	ADNI1
EcogPtMem	ADNI1
EcogPtLang	ADNI1
EcogPtVisspat	ADNI1
EcogPtPlan	ADNI1
EcogPtOrgan	ADNI1
EcogPtDivatt	ADNI1
EcogPtTotal	ADNI1
EcogSPMem	ADNI1
EcogSPLang	ADNI1
EcogSPVisspat	ADNI1
EcogSPPlan	ADNI1
EcogSPOrgan	ADNI1
EcogSPDivatt	ADNI1
EcogSPTotal	ADNI1
FLDSTRENG	ADNI3
DIGITSCOR_bl	ADNI2, ADNIGO, ADNI3
FLDSTRENG_bl	ADNI3
MOCA_bl	ADNI1
EcogPtMem_bl	ADNI1
EcogPtLang_bl	ADNI1

Variable name	Missing in cohort
EcogPtVisspat_bl	ADNI1
EcogPtPlan_bl	ADNI1
EcogPtOrgan_bl	ADNI1
EcogPtDivatt_bl	ADNI1
EcogPtTotal_bl	ADNI1
EcogSPMem_bl	ADNI1
EcogSPLang_bl	ADNI1
EcogSPVisspat_bl	ADNI1
EcogSPPPlan_bl	ADNI1
EcogSPOrgan_bl	ADNI1
EcogSPDivatt_bl	ADNI1
EcogSPTotal_bl	ADNI1
ABETA_bl	ADNI3
TAU_bl	ADNI3
PTAU_bl	ADNI3
PIB_bl	ADNI2, ADNIGO, ADNI3
AV45_bl	ADNI1
FBB_bl	ADNI1, ADNI2, ADNIGO

There are 33 subjects who change back to either CN or MCI after a Dementia diagnosis. All observations across all users account for 325 total observations. The distribution of the users may be seen in figure [A.1](#) and [A.2](#). It is important to notice that this is after the modifications made to the 'DX' column, as described in section [5.2.1](#).

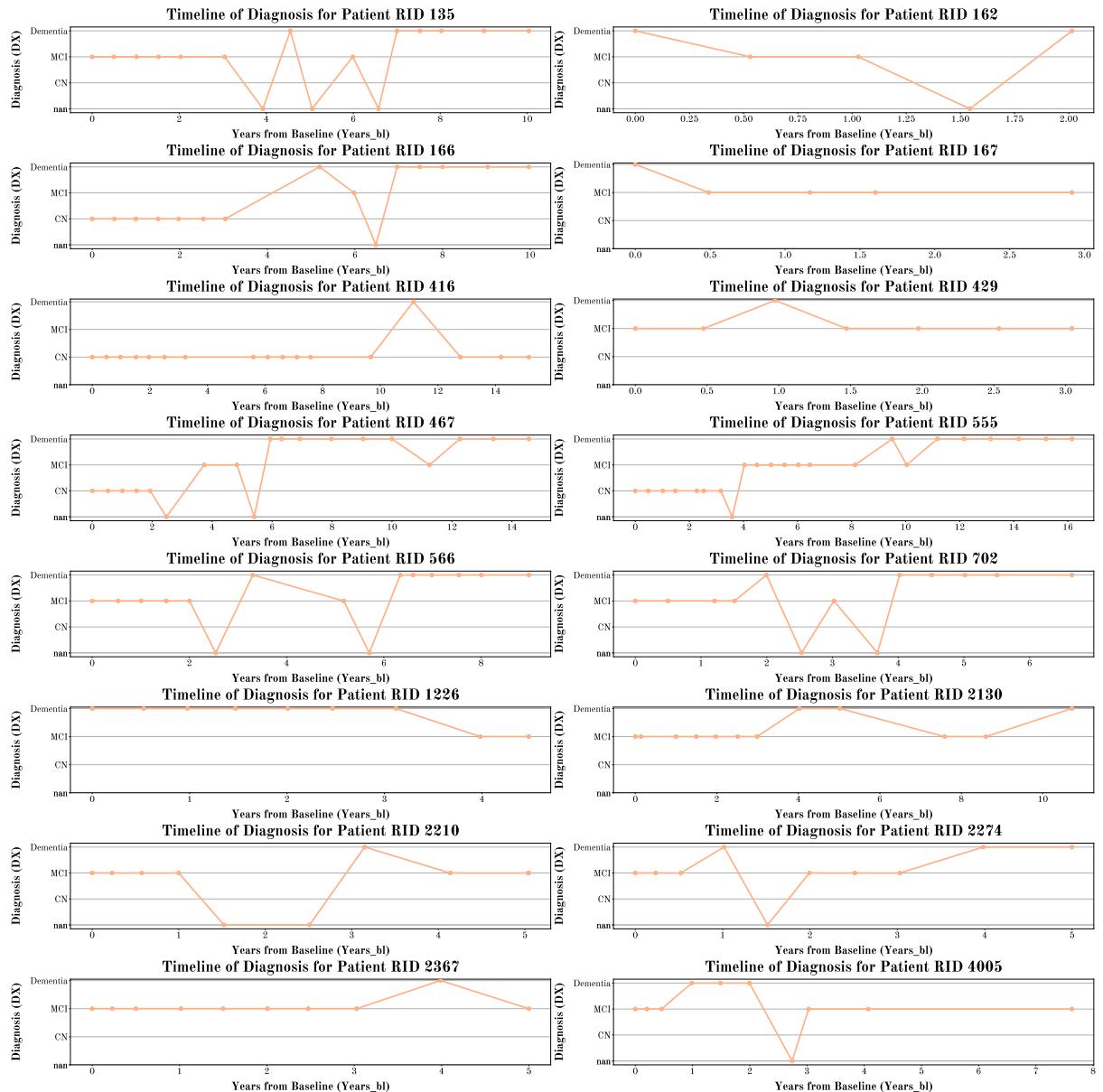


Figure A.1: Diagnosis over time for subjects who get diagnosed with AD and then back to another diagnosis.

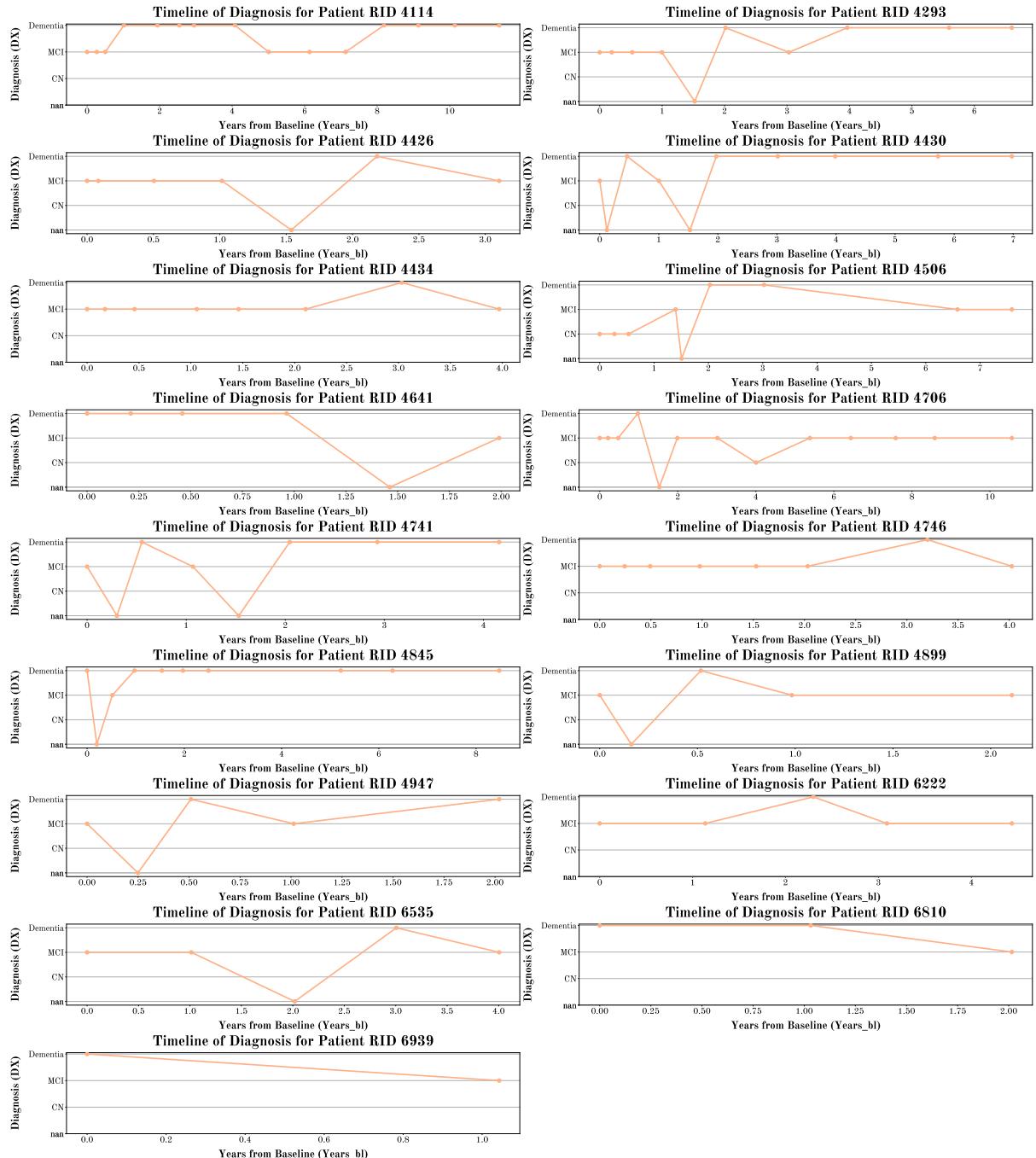


Figure A.2: Diagnosis over time for subjects who get diagnosed with AD and then back to another diagnosis.

APPENDIX B

MVAS Dataset

Modality	Variable	Datatype	Description
Demographical	mr_id	Integer	Subject ID.
	visit	Categorical	Visit, either bl or m24.
	sex_male	Boolean	True if male, false if female.
	age	Integer	Subject age at measurement.
	apoe4	Categorical	Number of APOE4 alleles, either 0, 1 or 2.
	Diagnosis	Categorical	Patient diagnosis, either healthy, mci or ad.
	mmse	Integer	cognitive assessment score.
	moca	Integer	Montreal Cognitive Assessment (MoCA).
	education_years	Integer	Years of education.
FTP SUVR	Entorhinal	Float	FTP SUVR values for the entorhinal region.
	Inferior_temporal		FTP SUVR values for the inferior temporal region.
	Meta_ROI	Float	Tau PET metaROI.
MRI imaging	param	Categorical	Indicates if all the subsequent brain regions are of type: Volume_mm3, PiB_SUVR, PK_bp, PWI_CBF, PWI_CBV, PWI_CTH, PWI_MTT, PWI_RTH, SEPWI_CBF, SEPWI_CBV, SEPWI_CMRO2, SEPWI_CTH, SEPWI_MTT, SEPWI_OEF, SEPWI_RTH, FTP_SUVR
	frontal_gm_left	Float	Measurement in left frontal grey matter.
	frontal_wm_left	Float	Measurement in left frontal white matter.
	frontal_gm_right	Float	Measurement in right frontal grey matter.
	frontal_wm_right	Float	Measurement in right frontal white matter.
	temporal_gm_left	Float	Measurement in left temporal grey matter.
	temporal_wm_left	Float	Measurement in left temporal white matter.
	temporal_gm_right	Float	Measurement in right temporal grey matter.
	temporal_wm_right	Float	Measurement in right temporal white matter.
	parietal_gm_left	Float	Measurement in left parietal grey matter.
	parietal_wm_left	Float	Measurement in left parietal white matter.

Modality	Variable	Datatype	Description
	parietal_gm_right	Float	Measurement in right parietal grey matter.
	parietal_wm_right	Float	Measurement in right parietal white matter.
	occipital_gm_left	Float	Measurement in left occipital grey matter.
	occipital_wm_left	Float	Measurement in left occipital white matter.
	occipital_gm_right	Float	Measurement in right occipital white matter.
	occipital_wm_right	Float	Measurement in right occipital grey matter.
	hippocampus_left	Float	Measurement in the left hippocampus.
	hippocampus_right	Float	Measurement in the right hippocampus.
	thalamus_left	Float	Measurement in the left thalamus.
	thalamus_right	Float	Measurement in the right thalamus.
	caudate_left	Float	Measurement in the left caudate.
	caudate_right	Float	Measurement in the right caudate.
	putamen_left	Float	Measurement in the left putamen.
	putamen_right	Float	Measurement in the right putamen.
	whole_brain	Float	Measurement in the whole brain.
	gray_matter	Float	Measurement in all grey matter.
	nawm	Float	Measurement in all white matter.

Table B.1: Overview and description of the complete MVAS dataset.