UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y DISEÑO INDUSTRIAL

Master in Industrial Engineering

MASTER'S THESIS

Estimates of the durations of the preclinical and prodromal stages of Alzheimer's disease using a data-driven approach

Amin Khatibi

Tutor: Carlos Platero

Departament: Industrial Engineering

Madrid, July, 2023

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y DISEÑO INDUSTRIAL

Master in Industrial Engineering

MASTER'S THESIS

Estimates of the durations of the preclinical and prodromal stages of Alzheimer's disease using a data-driven approach

Autor's Signature *Tutor's Signature*

Titel: Estimates of the durations of the preclinical and prodromal stages of Alzheimer's disease using a data-driven approach
Author: Amin Khatibi
Tutor: Carlos Platero

THE JURY

President:

Vocal:

Secretary:

Having carried out the defense and reading of the Final Degree Project on ..... of ..... of .... in ....., at the Escuela Técnica Superior de Ingeniería y Diseño Industrial of the Universidad Politécnica de Madrid, agrees to award the QUALIFICATION of:

VOCAL

SECRETARY PRESIDENT

Acknowledgements

I would like to express my sincere appreciation to everyone who helped me finish my thesis, "Estimates of the durations of the preclinical and prodromal stages of Alzheimer's disease using a data-driven approach."

First and foremost, I want to thank my supervisor Carlos Platero for their direction, encouragement, and insightful input throughout the project. His knowledge and skills in Alzheimer's disease and prediction models have been crucial in influencing the direction and scope of our effort.

I'm also thankful to the Alzheimer's Disease Neuroimaging Initiative (ADNI) personnel and researchers for granting me access to their huge database of individuals with amyloid pathology. This research would not have been achieved without their generosity.

I also would like to express my gratitude to the study participants whose clinical and imaging data were used in this investigation. Because of their willingness to participate in the ADNI project, researchers now have a significant resource for studying Alzheimer's disease and building prediction models.

Finally, I would like to express my gratitude to my friends and family for their constant support and encouragement during my studies. During difficult times, their love and encouragement kept me motivated and focused.

Summary

The goal of this thesis is to estimate the length of Alzheimer's disease's preclinical and prodromal phases using data from a database of individuals with amyloid pathology at various stages of cognitive impairment. To do this, the study will use a number of biomarkers, such as lumbar puncture markers, neuropsychological assessment scores, and clinical imaging, to develop prediction models and disease progression models utilizing Artificial Intelligence approaches. The research has four primary aims, which will be discussed in further depth. The first goal is to look at how Alzheimer's disease progresses in the ADNI cohort, from normal cognition to cognitive impairment and dementia. This will include following patients' cognitive impairment over time and detecting trends in clinical data that may suggest the development of Alzheimer's disease. The project will employ longitudinal data from the ADNI cohort, a major multi-center study of Alzheimer's patients, to do this.

The study will assess the tau pathology and neuro-degeneration profiles of the study individuals to reach the second goal. The buildup of tau protein in the brain, which is a characteristic of Alzheimer's disease, is referred to as tau pathology. Neuro-degeneration profiles, on the other hand, relate to changes in the brain that occur as the illness advances, such as brain cell loss. The project intends to acquire a better understanding of the underlying processes of Alzheimer's disease and uncover novel targets for treatments by studying these aspects. The third goal of the project is to use a longitudinal study to discover the biomarker predictors with the highest ability to estimate the progression of Alzheimer's disease. This will entail examining the links between various biomarkers and clinical outcomes in order to determine the most accurate predictors of disease progression. The study believes that by doing so, we may be able to better detect Alzheimer's disease in its early stages, when therapies are most successful.

Finally, the study intends to determine the duration of the disease's preclinical and prodromal stages based on the postulated natural history. The preclinical stage of Alzheimer's disease refers to the time before symptoms develop, whereas the prodromal stage refers to the time when symptoms appear but are not severe enough to fulfill the criteria for a dementia diagnosis. The study intends to acquire insights into the natural history of the disease and influence the development of new therapies and interventions by predicting the length of these stages.

Abstract

The goal is to quantify the time length of the preclinical and prodromal phases of Alzheimer's disease using a database of individuals with amyloid pathology and varying degrees of cognitive deterioration. We will begin with lumbar puncture markers, neuro-psychological measures scores, and other measures retrieved from the patients' clinical pictures. Artificial Intelligence approaches will be used to create prediction models and illness progression models using this data. The objectives will be broken down as follows:

1. to investigate the progression of Alzheimer's disease in the ADNI cohort of patients from normal cognition to cognitive impairment and dementia.
2. Examine the tau pathology and neurodegeneration characteristics of the research participants.
3. To identify the biomarkers with the best predictive potential for Alzheimer's disease progression through a longitudinal research.
4. Estimate the duration of the disease's preclinical and prodromal stages based on the postulated natural history.

Keywords: Alzheimer; Predictive Models; Machine Learning, Disease duration; Preclinical; Prodromal

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Project Motivation

Alzheimer's disease is a debilitating neurological illness that affects millions of individuals throughout the world. It is a progressive condition that starts with modest cognitive impairment and develops to dementia. While there is currently no cure for Alzheimer's disease, early identification and care can help to slow its course and improve patients' and their families' quality of life. One of the most difficult issues in Alzheimer's research is developing biomarkers that can properly predict the disease's start and progression. Lumbar puncture indicators, neuropsychological scores, and other clinical pictures from individuals with amyloid pathology are possible sources of biomarkers that may be utilized to develop prediction models and disease progression models.

The goal of this thesis research is to estimate the length of Alzheimer's disease's preclinical and prodromal phases in patients with amyloid pathology. To accomplish this purpose, the study will examine longitudinal data from patients in the ADNI cohort using Artificial Intelligence approaches. The research will specifically investigate the progression of Alzheimer's disease in patients from normal cognition to cognitive impairment and dementia, as well as identify the biomarker predictors with the best potential to estimate the progression of Alzheimer's disease. Furthermore, the project will examine the tau pathology and neuro-degeneration profiles of the study participants in order to better understand the underlying causes of Alzheimer's disease. By merging these many data sources, the research hopes to provide a full natural history of Alzheimer's disease that may be used to predict the duration of the illness's preclinical and prodromal stages.

Overall, the findings of this thesis study have the potential to further our understanding of Alzheimer's disease and give important insights into its early identification and intervention. Clinicians may be able to act early and maybe reduce the advancement of the disease by finding the biomarkers that are most predictive of disease progression, eventually improving the lives of patients and their families.

1.2 Objetives

1. Analyze the progression of Alzheimer's disease in the ADNI cohort participants from normal cognition to cognitive impairment and dementia.

2. Analyze the research subjects' tau pathology and neuro-degeneration profiles to better understand the underlying causes of Alzheimer's disease.

3. Through longitudinal research, identify the biomarker predictors with the highest ability to predict the progression of Alzheimer's disease.

4. Create prediction models and disease progression models utilizing Artificial Intelligence approaches to estimate the temporal length of Alzheimer's disease's preclinical and prodromal stages.

5. Create a detailed natural history of Alzheimer's disease that may be used to determine the duration of the disease's preclinical and prodromal stages.

1.3 Socio-Economic Considerations in Addressing Alzheimer's Disease: Protecting Participants' Rights and Promoting Equitable Care

Alzheimer's disease affects millions of people and their families, posing a serious threat to world health. Even though there is no cure, improvements in early identification and management might lead to better patient outcomes. This section examines the ethical issues surrounding research on Alzheimer's disease and evaluates the potential social, environmental, and financial repercussions of tackling this complicated problem.

Ethical considerations

1. Informed permission: Obtaining informed permission from study participants is necessary for Alzheimer's disease research including biomarker prediction models and disease progression analyses. Following ethical guidelines makes ensuring that participants are aware of the dangers, advantages, and potential uses of their data, promoting trust and respect for autonomy. [46]

2. Privacy and data protection: The use of AI algorithms and longitudinal data analysis raises questions about privacy and the protection of sensitive medical data. To ensure the confidentiality of participant data and prevent illegal access or abuse, ethical guidelines must be implemented. [44]

3. Equity and Access: It is crucial to provide equal access to early detection and care for people who are at risk of developing Alzheimer's disease or who already have it. Important ethical issues include addressing socioeconomic inequalities and advancing inclusion in research involvement and healthcare services.

Social Impact

1. Enhanced Quality of Life: Alzheimer's disease may be quickly identified, and early intervention based on reliable prediction models has the potential to greatly enhance patients' and their families' quality of life. Early assistance and interventions can lighten the load on caregivers and help with better planning for future care requirements.

2. Education and Awareness: Research projects on Alzheimer's disease work to increase public understanding of the disease, its risk factors, and potential

preventative strategies. Increased information enables people to make wise health decisions and inspires proactive involvement in promoting brain health. [25]

3. Reduction of Stigma: Research initiatives can work to combat the stigmatization connected with Alzheimer's by improving our understanding of the illness. A more positive public image encourages compassion for those who are affected and encourages others to assist them, which increases societal acceptability.

Economic Impact

1. Savings on costs: Early detection and treatments may lessen the financial burden associated with Alzheimer's disease. Healthcare expenses associated with hospitalizations, long-term care, and caregiver assistance can be reduced by decreasing disease development and postponing severe cognitive loss.

2. Research and Innovation: New discoveries in treatment methods, preventative measures, and better healthcare delivery may be made as a result of advances in Alzheimer's disease research. This may have favorable economic effects, such as the creation of jobs, funding of research infrastructure, and increased competitiveness in the pharmaceutical and biotechnology industries. [60]

1.4 Document structure

In order to facilitate the reading of the document, the content of each chapter is detailed below.

- Chapter 1: An introduction is given.
- Chapter 2: The articles that have been taken as a starting point for the development of the project are presented.
- Chapter 3: The different materials used in the work are presented.
- Chapter 4: The methods of the different parts of the study are shown, with a primary focus on the coding.
- Chapter 5: The results of the different parts of the study are shown.
- Chapter 6: The discussions obtained from the results are presented.
- Chapter 7: The final conclusions of the project are presented.
- Appendices: Additional information is provided extending that of the main work.

Chapter 2

Literature Review

In this chapter the used reference articles will be summarised including the relevant plots, graphs and formula. The chapter will be ended with the applications of each article on this paper.

2.1 NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease (Jack et. all 2018) [34]

2.1.1 Introduction

The article emphasizes the significance of using biological indicators to define Alzheimer's disease (AD) rather than only relying on clinical symptoms. In the past, AD was usually identified using a combination of symptoms and an autopsy (post-mortem investigation). The line separating clinical symptoms from neuropathological alterations has, however, faded over time. As a result, the authors suggest an AD research strategy built on a biomarker-based definition. This framework is focused on identifying people who have certain biomarkers, such as amyloid plaques and tau tangles, which are recognized neuropathological abnormalities linked to AD. The framework does, however, offer a way for incorporating these people in the research together with others who are on the Alzheimer's continuum, even if not all participants may be able to be categorized by these biomarkers. This biomarker-based approach would bring all medical professionals and researchers together and increase our knowledge of Alzheimer's disease and its treatment.

2.1.2 AD Defnition, Biomarkers and AT(N) Classification

The committee described Alzheimer's disease (AD) as a biologic entity that may be recognized in live individuals by biomarkers. The biomarkers that were taken into consideration to characterize AD were those exclusive to the characteristic proteinopathies of AD, namely pathologic tau and Ab (b-amyloid). According to studies, those who have aberrant amyloid biomarkers typically experience the illness more quickly and these biomarkers are the first signs of AD neuropathologic change that can now be seen in live people. The committee came to the conclusion that aberrant b-amyloidosis biomarkers might act as the only indicator of the pathologic alteration associated with Alzheimer's. But for the condition to be referred to as "Alzheimer's disease," there must be proof of both pathologic tau abnormalities

and Ab abnormalities. Alzheimer's pathologic change would be given to people who just had biomarker evidence of Ab deposition. Alzheimer's disease and Alzheimer's pathologic change aren't thought of as independent things, but rather as the early and later stages of the "Alzheimer's continuum." Based on their biomarker profiles, each person may be classified into one of three main biomarker categories: those with normal AD biomarkers, those with non-AD pathologic change, or those who are on the Alzheimer's continuum. Either Alzheimer's pathologic change or Alzheimer's disease are referred to as the "Alzheimer's continuum". See table 2.1.

Research on Alzheimer's disease (AD) and brain aging frequently makes use of several imaging techniques and cerebrospinal fluid (CSF) biomarkers. A methodical strategy was used to guarantee a framework for the research that can be applied generally. A '+' denotes abnormal levels and a '-' denotes normal levels in this classification system, which is called AT(N), which divides the biomarkers into three groups depending on the types of pathological processes that each one measures:

1. **A** comprises biomarkers of Ab plaques, such as cortical amyloid PET ligand binding or low CSF Ab42.
2. **T** encompasses biomarkers of fibrillar tau, such as elevated CSF phosphorylated tau (P-tau) and cortical tau PET ligand binding.
3. **(N)** represents biomarkers of neurodegeneration or neuronal injury, including cerebrospinal fluid T-tau, FDG PET hypometabolism, and atrophy on MRI.

Cerebrospinal fluid (CSF) biomarkers and imaging biomarkers are both used in the diagnosis of AD. The CSF biomarkers assess the concentration of proteins within the cerebrospinal fluid at a particular time, which offers insight into both the production and elimination of proteins, while the imaging biomarkers assess the amount of damage that has accumulated within the brain over time. Given that AD is a chronic illness that can endure for many years or even decades, there may be differences between the results from these two biomarkers, but they ultimately coincide in the long term.

Tau PET, a more recent imaging biomarker, is being used to diagnose AD and is thought to be a trustworthy indication of tau accumulation. Two CSF biomarkers, T-tau and P-tau, are used to diagnose AD, with P-tau being the most well studied of the two. However, there is a significant association between T-tau and P-tau in AD patients. P-tau is the only biomarker that has consistently shown an increase in AD patients, despite the fact that these biomarkers may behave similarly in AD but differently in other disorders.

2.1.3 Cognitive stages

Both interventional trials and observational research investigations can make use of the framework. In observational research, all recruited sample members are included regardless of their biomarker status, in contrast to interventional studies where participants are often chosen based on the existence of specific biomarkers. The NIA-AA framework specifies two categorical clinical staging approaches to suit these various demands. The first system classifies the cognitive continuum into three levels using the syndromal category staging scheme:

1. Cognitively Unimpaired (CU)

<table>
<tr><th>AT(N) profiles</th><th colspan="2">Biomarker category</th></tr>
<tr><td>A-T-(N)-</td><td colspan="2">Normal AD biomarkers</td></tr>
<tr><td>A+T-(N)-</td><td>Alzheimer's pathologic change</td><td rowspan="4">Alzheimer's continuum</td></tr>
<tr><td>A+T+(N)-</td><td>Alzheimer's disease</td></tr>
<tr><td>A+T+(N)+</td><td>Alzheimer's disease</td></tr>
<tr><td>A+T-(N)+</td><td>Alzheimer's and concomitant suspected non Alzheimer's pathologic change</td></tr>
<tr><td>A-T+(N)-</td><td colspan="2">Non-AD pathologic change</td></tr>
<tr><td>A-T-(N)+</td><td colspan="2">Non-AD pathologic change</td></tr>
<tr><td>A-T+(N)+</td><td colspan="2">Non-AD pathologic change</td></tr>
</table>

Table 2.1: AT(N) Classification
[32]

2. Mild Cognitive Impairment (MCI)

3. Dementia

The recruited cohort as a whole, including individuals with normal Alzheimer's biomarkers and non-Alzheimer's pathology as well as those in the Alzheimer's continuum, are all subject to this plan. See figure 2.2. Only individuals who fall inside the Alzheimer's continuum can employ the second type of staging system, the numeric clinical staging scheme, which may prove particularly helpful in clinical studies. Additionally, the NIA-AA paradigm recognizes the value of cognitive staging both in situations where participants have had access to earlier longitudinal clinical or cognitive testing evaluations and in situations where they have not and are being examined for the first time. The framework offers guidelines for integrating biomarker profiles with cognitive stages and coined the term "prodromal Alzheimer's disease" to designate people with an early-stage Alzheimer's biomarker profile.

The numerical clinical staging is as follows:

1. Cognitive test results are within the predicted range. There have been no reports of cognitive deterioration or a new development of neurobehavioral disorders. No signs of deterioration or symptoms were noted by the observer or, if available, by longitudinal cognitive testing.

2. Cognitive test results reveal typical performance. Any cognitive region, though, may see a reduction in prior cognitive capability. The subject may experience a reduction in cognitive function, or over time, minor changes in cognitive tests may show a deterioration. Mild neurobehavioral changes like those in motivation, anxiety, or mood may also occur in certain people. These symptoms

Biomarker Profile		Cognitively unimpaired	MCI	dementia
	$A^-T^-(N)^-$	normal AD biomarkers, cognitively unimpaired	normal AD biomarkers with MCI	normal AD biomarkers with dementia
	$A^+T^-(N)^-$	Preclinical Alzheimer's pathologic change	Alzheimer's pathologic change with MCI	Alzheimer's pathologic change with dementia
	$A^+T^-(N)^+$	Alzheimer's and concomitant suspected non Alzheimer's pathologic change, cognitively unimpaired	Alzheimer's and concomitant suspected non Alzheimer's pathologic change with MCI	Alzheimer's and concomitant suspected non Alzheimer's pathologic change with dementia
	$A^+T^+(N)^-$	Preclinical Alzheimer's disease	Alzheimer's disease with MCI (Prodromal AD)	Alzheimer's disease with dementia
	$A^+T^+(N)^+$			

Non-Alzheimer's continuum profiles are not included in table because the risk associated with different combinations of T+(N)-, T+(N)+, T-(N)+ among A- individuals has not been established

□ rate of short term clinical progression expected to be low

■ rate of short term clinical progression expected to be high

Table 2.2: Syndromal Cognitive Stage
[33]

must be recent in start, persistent, and unrelated to current events. Daily activities are unaffected despite the reduction.

3. When tested longitudinally, subjects with aberrant cognitive test scores exhibit decrease from baseline, as stated by the subject or an observer. In addition to memory, decline may affect other cognitive processes as well. Daily tasks are still carried out autonomously, but with some little difficulty when they are complicated, according to self-report or research partner verification.

4. Deterioration in various cognitive areas and/or neurobehavioral disruption are symptoms of mild dementia. This can be mentioned by the subject, by a third party, or by looking at the outcomes of cognitive tests. The decrease impairs routine tasks of daily living, primarily instrumental ones, necessitating sporadic help and reducing complete independence.

5. Progressive Cognitive Impairment, a reduction in cognitive function, and/or modifications in neurobehavioral skills are all symptoms of moderate dementia. This has a significant functional influence on everyday activities, such as making it difficult to do simple chores, need frequent help, and causing loss of independence.

6. A clinical interview may be impossible in cases of severe dementia due to progressive cognitive impairment, a reduction in cognitive function, or neurobehavioral abnormalities. due to the substantial impact on fundamental daily tasks and self-care, full reliance results.

2.2 Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype (Vermunt et. all 2019) [58]

2.2.1 Introduction and methods

This study aimed to determine the length of each stage of Alzheimer's disease by using a multistate modeling approach, which has been previously used in other AD studies. The researchers of this study respected a couple of factors such as age, the setting in which the patient is examined (clinical vs. research), sex (male or female), APOE genotype, and baseline CSF tau levels. The data was collected from six longitudinal cohort studies and the participants were selected according to following criteria: 50 years or older and also showing evidence of amyloid accumulation. AD was divided into four stages: preclinical AD, prodromal AD, mild AD dementia, and moderate-to-severe AD dementia. Besides collecting mortality data, the study also took in account predictor variables such as age, sex, and setting in the analysis. The multistate model was employed to approximate the length of the disease and the impact of different covariates on mortality. This model was fitted with distinct numbers of covariates, for example: setting, sex, APOE genotype, and tau, with the final model containing all five covariates. The results of this study provide valuable insights into the duration of AD stages and the impact of various factors on mortality.

2.2.2 Results and discussion

The study investigated the progression of Alzheimer's disease (AD) by analyzing data from six different groups of participants. The analysis included a total of 3268 people with an average age of 73 years at the start of the study. The length of the study varied from 0.3 to 20 years, with an average follow-up time of 2.8 years. Results showed that 32% of participants progressed to at least one subsequent stage of AD. See table 2.3. Factors such as age, except for mortality in the preclinical AD stage and progression from prodromal AD to mild AD dementia, influenced the transition rates to subsequent stages of AD. It was also observed that a higher progression rate was recorded in a clinical setting compared to a research setting. Women showed a higher progression rate from mild AD to moderate AD dementia and had a lower risk of mortality in moderate AD dementia. Carriers of APOE $\epsilon 4$ had a higher rate of progression from preclinical AD to prodromal AD and from prodromal AD to mild AD dementia. The duration of the moderate AD dementia stage was longer in women than in men. The overall predicted disease duration was influenced by factors such as APOE genotype, age, and setting.

The results indicate that age has the strongest effect on the duration of the preclinical and dementia stages, with higher progression and mortality rates. The effect of setting showed that the preclinical and prodromal stage was shorter in the clinical setting compared to the research setting. APOE genotype showed a shorter age-specific duration of the preclinical stage in APOE $\epsilon 4$ carriers. Females had a lengthier dementia stage duration amid the lower mortality. The presence of increased CSF t-tau was associated with a shorter predementia disease period. In some cases, the total estimated duration of the disease was longer than the

Characteristics	Preclinical AD (n = 438)	Prodromal AD (n = 729)	Mild AD dementia (n = 1867)	Moderate-to-severe AD dementia (n = 234)	*P* value overall group difference
Age (years)	73 (7)	72 (7)	73 (9)	75 (10)	<0.01*
Male (n)	204 (47%)	417 (57%)	781 (42%)	74 (33%)	<0.01
MMSE [0-30, median (interquartile range)] (n = 3252)	29 (28-30)	27 (26-29)	22 (19-24)	16 (13-19)	<0.01†
APOE ε4 allele carrier‡ (n) (n = 1984)	210 (49%)	466 (66%)	554 (71%)	35 (51%)	<0.01
Abnormal CSF total tau§ (n) (n = 1563)	87 (38%)	346 (57%)	535 (80%)	47 (82%)	<0.01
Follow-up years [median (interquartile range)]	3.8 (2-4.5)	3.9 (2.5–4.8)	2.0 (1.5–2.5)	2.0 (1.2–2.3)	<0.01†
Progression to next clinical disease stage (n)	87 (20%)	325 (45%)	569 (30%)	NA	NA
Death at follow-up (n)	12 (3%)	76 (10%)	215 (12%)	54 (23%)	NA
Participants by cohort (n, ADC/ADNI/AIBL/DESCRIPA/ Gothenburg/ICTUS)	40/180/191/23/4/0	140/449/73/49/18/0	507/224/69/0/1/1066	64/1/3/0/0/166	NA

Table 2.3: Baseline characteristics
[57]

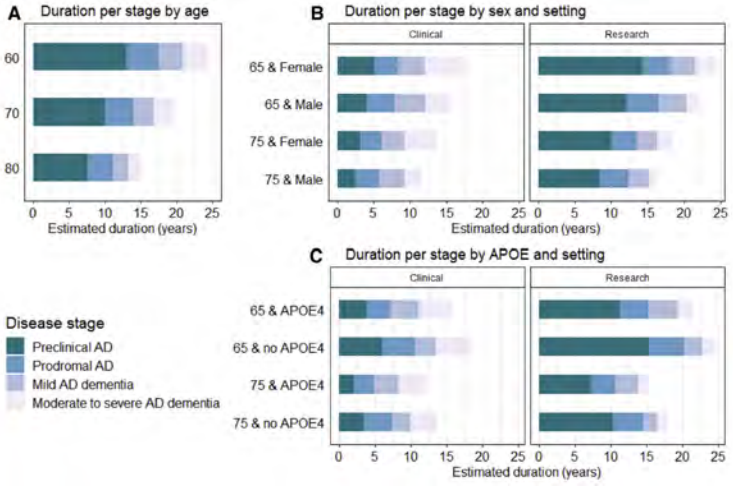

Figure 2.1: Results
[59]

residual life expectancies, which may have been caused by participants who were healthier or by an overestimation.. The MSM approach is a strength of the study, but it has limitations, such as combining data from multiple cohorts and not being representative of the general population. Clinicians can use these findings to provide patients and their families with prognostic information.

2.3 CSF biomarkers of Alzheimer's disease concord with amyloid-β PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts (Hannson et al. 2018) [27]

2.3.1 Introduction and Methods

Treatments for Alzheimer's disease (AD) are at the moment limited to providing only temporary relief, however, researchers are working to develop medical drugs that can modify the disease. An accurate diagnosis of AD is decisive for effective treatment, although the current method based on clinical symptoms is not always reliable. The usage of biomarkers like amyloid-β and tau protein is endorsed to

optimize the accuracy of diagnosis. The only FDA-approved approach is a visual analysis of amyloid-β PET scans, but this method is costly and needs specialized equipment. Cerebrospinal fluid biomarkers such as Aβ(1-42), pTau, and tTau can divide AD patients from those without the disease and could also predict future progression. However, the available CSF assays for these biomarkers vary largely between laboratories and batches. Roche Diagnostics is developing Elecsys CSF immunoassays that are notably precise and consistent. A study was conducted to determine if these assays can be utilised to establish global cutoffs that can be applied to different populations, even when analyzed in distinct laboratories. The conclusion was made that the Elecsys CSF immunoassays were consistent with amyloid PET classification in two patient groups and had the ability to predict clinical progression in patients with mild cognitive impairment.

The scientists followed a three-step methodology to find out if CSF biomarkers could predict how well someone was doing on a PET scan measuring the presence of amyloid-β proteins in the brain. Then, they made use of a "CSF cutoff adjustment factor" to adjust those biomarker results so that they could be used to predict future clinical progression. Finally, they tested the ability of the adjusted cutoff to predict future clinical progression in a second, different and independent group.

2.3.2 Study population

The study was done on two separate populations and containts three parts. In the first part, the BioFINDER study, 728 patients were called up between 2010 and 2014 (with normal controls, with mild cognitive impairment, with Alzheimer's disease). The primary focus was to resolve the correlation between cerebrospinal fluid (CSF) biomarkers and amyloid-b PET imaging by analysing the first group of 277 patients with mild cognitive impairment. The second part of the study looked at differences of the effects of the use of the two different protocols (BioFINDER and ADNI) for conducting CSF samples taken from 20 subjects who endured diagnostic lumbar puncture. The results of this comparison were used to establish a "CSF cut-off adjustment factor." The last part of the study was the ADNI validation study. This study included 918 subjects (cognitively normal, mild cognitive impairment, Alzheimer's disease). The primary analysis group of 646 participants was utilised to validate the concordance of the CSF biomarker cutoffs with the amyloid-β PET classification and to assess the ability of CSF biomarker status to predict future clinical progression in Alzheimer's disease.

This paper focuses on studies of predictability of clinical progression, visualization of CSF deposition and evaluation of CSF biomarkers in AD. The clinical dementia rating–sum of boxes (CDR-SB) scores of 619 subjects with early or late mild cognitive impairment (MCI) at baseline were followed for 2 years using ADNI data. Cerebral amyloid-β deposits were visualized with positron emission tomography (PET) imaging using either the PET tracer [18F]flutemetamol (BioFINDER) or [18F]florbetapir (ADNI). The images were reevaluated by three independent readers and quantitatively assessed using the standardized uptake value ratio (SUVR). CSF samples were collected and biomarkers such as β-amyloid(1-42), phosphorylated-tau (181P) and total-tau (Ttau) were measured. CSF biomarker cutoffs were determined to optimize match with BioFINDER's visual-read PET classification based on performance and robustness. Pre-analytical procedures for CSF samples were compared

Cohort	CSF biomarker	Cutoff	PPA, %	NPA, %	OPA, %	AUC, %
BioFINDER	Aβ(1–42)	1100 pg/mL	90.9 (83.9–95.6)	72.5 (65.0–79.1)	79.8 (74.6–84.4)	86.5 (82.3–90.7)
	pTau/Aβ(1–42)	0.022	90.9 (83.9–95.6)	89.2 (83.5–93.5)	89.9 (85.7–93.2)	94.4 (91.5–97.3)
	tTau/Aβ(1–42)	0.26	90.9 (83.9–95.6)	89.2 (83.5–93.5)	89.9 (85.7–93.2)	94.0 (91.0–97.0)
ADNI	Aβ(1–42)	880 pg/mL	83.6 (79.3–87.3)	85.3 (80.8–89.1)	84.4 (81.3–87.1)	92.1 (90.0–94.3)
	pTau/Aβ(1–42)	0.028	88.2 (84.3–91.4)	92.6 (89.1–95.3)	90.3 (87.7–92.4)	96.3 (95.2–98.0)
	tTau/Aβ(1–42)	0.33	85.0 (80.8–88.6)	94.0 (90.7–96.4)	89.2 (86.5–91.5)	96.3 (94.8–97.7)

Table 2.4: Performance of CSF biomarker cutoffs [29]

between the BioFINDER and ADNI cohorts. Cut-off performance was assessed by assessing the agreement of CSF biomarkers with PET visual read-based and SUVR-based classifications. There was made use of a linear mixed-effects model to analyze the predictive properties of the CSF biomarkers. This model was adjusted for demographic variables such as age, gender, education, CDR-SB score at baseline, and APOE genotype. The results of this study demonstrated that CSF biomarker cutoffs were successfully matched optimally with BioFINDER's visual-read PET classification and that pre-analytical processing procedures were comparable between the BioFINDER and ADNI cohorts. CSF biomarkers were found to have predictive properties for clinical progression in AD.

2.3.3 Results

In the first part of the study the aim was to determine appropriate cutoff levels for biomarkers in cerebrospinal fluid (CSF) that are concordant with visual read classifications of amyloid-b PET scans. The cohort consisted of 277 individuals, with 110 (40%) being positive and 167 (60%) negative according to the majority vote of three independent visual reads. The interreader agreement was high with an average of 90.1%. The distribution of CSF biomarker concentrations was found to correspond with the PET read classifications. Cutoffs for Aβ(1–42), pTau/Aβ(1–42), and tTau/Aβ(1–42) were chosen based on values that separated the PET-positive and PET-negative groups effectively and were stable across measurement levels. The cutoff for Aβ(1–42) was set at 1100 pg/mL, while the cutoffs for pTau/Aβ(1–42) and tTau/Aβ(1–42) were set at 0.022 and 0.26, respectively. The distribution of CSF levels of pTau and tTau versus Aβ(1–42) revealed two clusters that corresponded with the PET classification. The use of pTau/Aβ(1–42) and tTau/Aβ(1–42) cutoffs showed a higher negative predictive accuracy (89%) than Aβ(1–42) alone (73%) with the same positive predictive accuracy (91%) resulting in an overall accuracy of 90%. There was a strong correlation between CSF pTau and tTau measurements, and there was no clear preference for either tau biomarker. See table 2.4.

In the second part, the authors examined the differences in levels of cerebrospinal fluid biomarkers (Aβ(1-42), pTau and tTau) between two different preanalytical protocols (BioFINDER and ADNI) on samples from the same patient. I checked. They found systematic differences (24%) in the measured levels of Aβ(1-42) in the cerebrospinal fluid, but did not find significant differences in the measured levels of pTau and tTau. To account for these differences, the authors apply threshold correction factors of 0.8 for Aβ(1–42) and 0.821 for pTau/Aβ(1–42) and tTau/Aβ(1–42). We calculated and validated these adjusted thresholds in Part 3. research. The final cut-off values were defined as Aβ(1-42) = 880 pg/mL, pTau/Aβ(1-42) = 0.028 and tTau/Aβ(1-42) = 0.33.

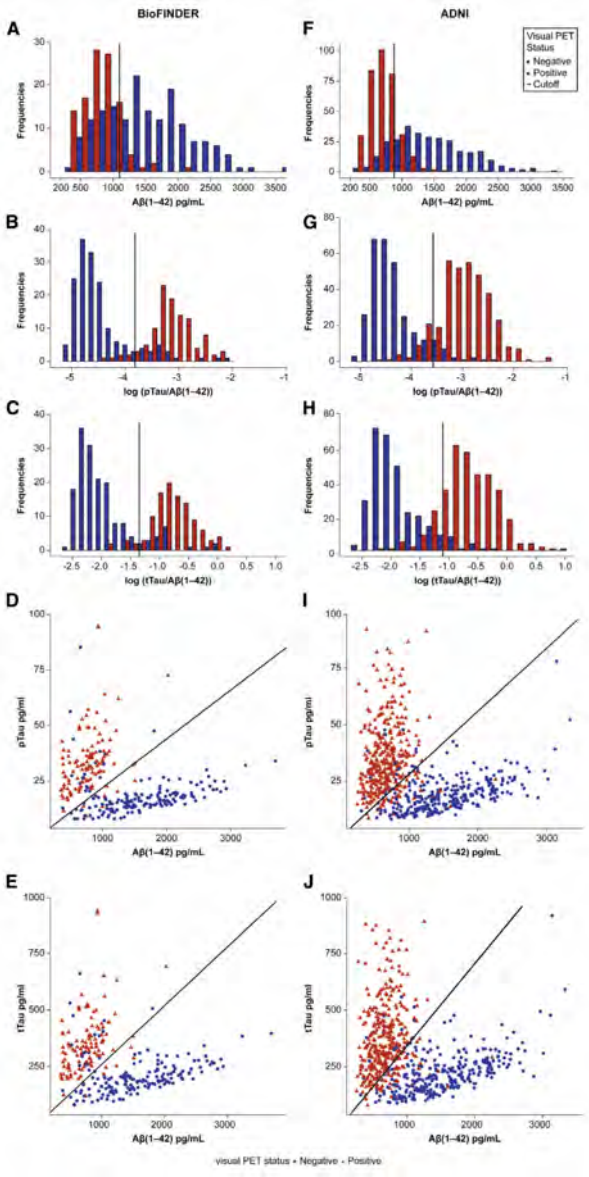

Figure 2.2: Distribution of the CSF biomarkers colored by PET visual read classification. (A–C) (BioFINDER cohort) and (F–H) (ADNI cohort): Frequency distribution of Aβ(1–42), log(pTau/Aβ[1–42]) and log(tTau/Aβ[1–42]), respectively, by PET classification. (D and E) (BioFINDER cohort) and (I and J) (ADNI cohort): Scatterplots of Aβ(1–42) versus pTau (D and I) and tTau (E and J) with the cutoffs for the respective ratio pTau/Aβ(1–42) (BioFINDER: 0.022, ADNI: 0.028) and tTau/Aβ(1–42) (BioFINDER: 0.26, ADNI: 0.33) shown as diagonal lines. n = 277 (BioFINDER A–E) and n = 646 (ADNI, F–J). Red bars or triangles, PET-positive; blue bars or dots, PET-negative.

[28]

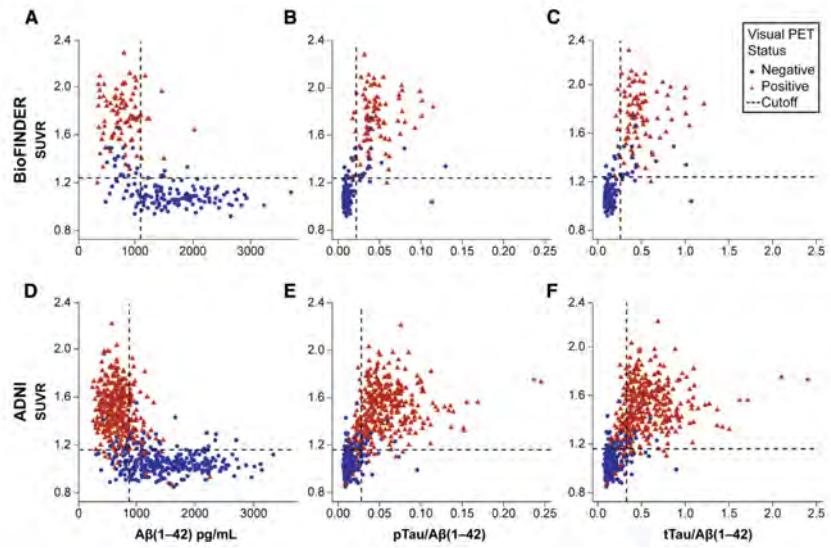

Figure 2.3: Scatterplots of SUVRs vs CSF biomarkers in BioFINDER (A–C) and ADNI [30]

Quantitative SUVR values were compared with qualitative imaging data from amyloid-β PET scans. The results showed high concordance between the two methods in the BioFINDER and ADNI groups. This study also found high concordance between CSF biomarkers and SUVR-based classification, with slightly higher concordance in the BioFINDER cohort than in the ADNI cohort. Agreement between CSF biomarkers and SUVR-based classification was very high for pTau/Aβ(1-42) and tTau/Aβ(1-42) ratios. The ADNI MCI population (n = 619) was examined to see if predefined cut-offs for CSF biomarkers could predict clinical development. The results indicate that there was a significant difference in progression between biomarker-positive subjects and biomarker-negative subjects (as measured by CDR-SB change), with biomarker-positives progressing more. See figure 2.4 This trend was evaluated for all three CSF biomarkers, pTau/Aβ(1-42) and tTau/Aβ(1-42) ratios were more predictive of progression between the two groups than Aβ(1-42) shows a significant difference. Results remained significant even after adjustments of the APOE ϵ4 status.

2.4 Categorical predictive and disease progression modeling in the early stage of Alzheimer's disease (Platero and Tobar 2022) [47]

2.4.1 Introduction and Materials

Many dementia therapies target the preclinical period, giving the chance to stop or postpone the emergence of symptoms. Examining patterns of change in a variety of indicators, such as cognitive functioning, MRI scans, and cerebrospinal fluid biomarkers, is required to comprehend the development of Alzheimer's from the early, asymptomatic stage to late-stage dementia. Cognitive tests are the conven-

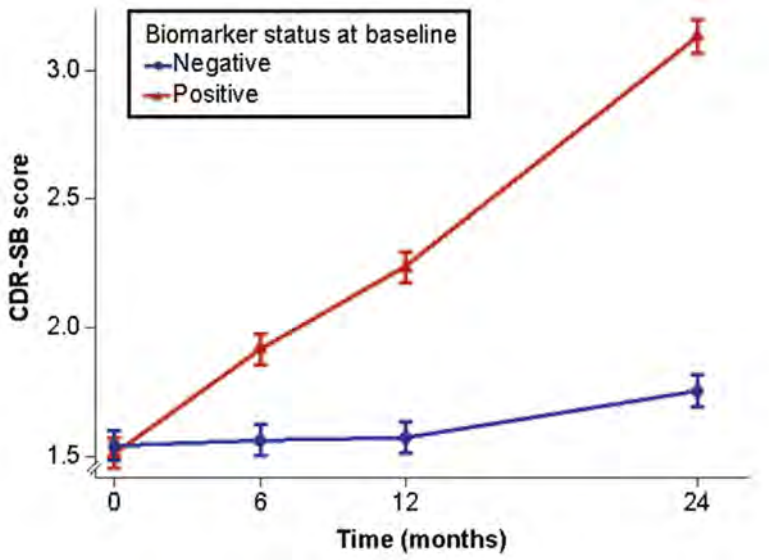

Figure 2.4: Time course of pTau/Aβ
[30]

tional way for diagnosing cognitive impairment, moderate cognitive impairment, or dementia; however, a more precise prediction may be produced by considering numerous domains. Cognitive deterioration and biomarker progression are increasingly seen as continuing biological processes that take place over an extended length of time in Alzheimer's disease.

Age, gender, and APOE status are only a few of the variables that affect the risk of progression. Combining factors associated to symptom onset, such as cerebrospinal fluid biomarkers, brain volume, and cognitive test scores, has been attempted to enhance the prediction of Alzheimer's risk. In certain research, automated methods have been utilized to categorize and forecast the diagnosis utilizing information from various sources. In order to identify unique patterns of change and provide details for a differential diagnosis, longitudinal studies are essential. A machine learning system has recently been created to view the course of Alzheimer's disease as a continuous process, extracting long-term pathological trajectories from various data sources. By combining temporal modeling of biomarker trajectories and extending Cox survival analysis, survival models that take into consideration a particular clinical group and account for conversion times, finite follow-up, or censoring may be produced. This makes it possible to include exploratory variables that evolve over time. The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset was used to examine a two-stage technique for describing the course of Alzheimer's disease. 56 sites in the United States and Canada were used to gather more than 2100 participants, ages 54 to 92. For as long as 10 years, the study has been observed. ADNI has been used in several articles to study age-related brain changes and the early detection of AD. The ADNI has established procedures for contrasting outcomes from various locations. The main goals of the tests were to assess cognitive function, brain imaging, and CSF levels of beta-amyloid, tau, and phosphorylated tau. The subjects of a second investigation were patients who had beta-amyloid data. It was possible to anticipate cognitive decline by keeping track of patients who were first diagnosed with cognitive impairment (CU) and identifying those whose symptoms

had advanced to MCI or dementia over time. The conversion time was noted for those who showed advancement.

2.4.2 Methods

Survival Analysis and Feature Selection

P markers were measured from n individuals over a series of follow-up times. The outcome for the kth marker of the individual i at time j was recorded as y_{ijk}, with i begin equal to 1...n, k equal to 1...p, and j equal to 1...q_{ik}. The data was analyzed using a Longitudinal Mixed Effects (LME) model, this was mathematically expressed as:

$$y_{ijk} = x' t_{ijk} \beta_k + \alpha_{0ik} + \alpha_{1ik} t_{ijk} + e_{ijk}$$

with t_{ijk} equal to a short-term observation time, x'_{ijk} equal to the row vector for the fixed effects (including age and scan time) and β_k equal to fixed effect coefficients. The random intercept and slope for each subject and outcome are represented by α_{0ik} and α_{1ik} respectively. The combination of these two values, represented by the vector $(\alpha_{0ik}, \alpha_{1ik})$, follows a bivariate Gaussian distribution with a mean of zero and a covariance matrix $\sum k$. This represents how the regression parameters for the i-th subject deviate from the population parameters. Additionally, eijk represents a measurement error that follows a Gaussian distribution with zero mean and a variance of σk.

A number of longitudinal measures taken from a population of CU patients were analyzed using the Longitudinal Mixed Effects (LME) model. This allowed for the approximate estimation of each marker's value for each subject over time. Additionally, it was determined if the participants had dementia or MCI throughout the observational period. When calculating the conversion time for pCU participants, the baseline was utilized, but for sCU individuals, the censoring time was used. Each important discrete period was given its own extended Cox model, which may take both independent and dependent factors into account throughout time.

$$h_{ij} = h_j \exp\left(\sum_{k=1}^{p_1} \eta_k \cdot y_{ijk} + \sum_{l=1}^{p_2} \theta_l \cdot z_{il}\right)$$

The hazard ratio (HR) calculates the relative risk of an event occurring for one individual compared to another one w.r.t. their specific time-varying characteristics such as y_{ijk}, z_{il}. This is further calculated using the baseline hazard function (h_j) and the effects of p1 and p2, which are independent variables y_{ijk}, z_{il} with corresponding coefficients ηk, ωl. The HR is determined by comparing the risk of an individual with characteristics y_{ijk}, z_{il} to that of a reference individual with characteristics y_{ijk}, z_{il} at the same point in time.

$$HR_{irj} = \frac{h_{ij}}{h_{rj}} = \exp\left(\sum_{k=1}^{p_1} \eta_k \left(y_{ijk} - y_{rjk}\right) + \sum_{l=1}^{p_2} \theta_l \left(z_{il} - z_{rl}\right)\right)$$

The hazard ratio, written as HR, is a measure of the differential risk of disease conversion between two subjects, one being characterized by the variables y_{ijk} and z_{il} and the other one being the reference subject characterized by y_{rjk} and z_{rl}. If $HR > 1$, the subject characterized by y_{ijk} and z_{il} has an increased risk of disease

conversion compared to the reference subject, whereas if $HR < 1$, the risk of disease conversion is decreased.

In the study, three models—one at the beginning of the trial and the other two for follow-up at 12 and 24 months—were constructed to predict the conversion of Cognitively Unimpaired (CU) individuals to Mild Cognitive Impairment (MCI) or dementia. For each of these models, the hazard ratios were determined and transformed into the likelihood that CU would progress to MCI or dementia using a logistic regression model.

$$p_{irj} = \frac{1}{1 + \frac{1}{HR_{irj}}}$$

Let HR_{irj} denote the hazard ratio at visit j, where $j \in (0, 12, 24)$. The vectors of exploratory variables for the subject and the reference at visit j are denoted by $\mathbf{y}_{ijk}$ and $\mathbf{z}_{il}$, and $\mathbf{y}_{rjk}$ and $\mathbf{z}_{rl}$, respectively. These vectors are formed with p_1 time-varying and p_2 time-independent variables, the latter modeled using linear mixed-effects (LME) models. HR_{irj} was estimated using an extended Cox-LME model with $\mathbf{y}_{rjk}$ and $\mathbf{z}_{rl}$, calculated using a random subset of the training population at visit j, which was sampled to include the same number of subjects representing both the stable cognitively unimpaired (sCU) and prodromal cognitively unimpaired (pCU) patients. The components of $\mathbf{y}_{rjk}$ and $\mathbf{z}_{rl}$ were defined by the average values of this population and scaled by their standard deviations, resulting in each exploratory variable being defined as a z-score. If $\mathrm{HR}_{irj} > 1$ for a subject with $\mathbf{y}_{ijk}$ and $\mathbf{z}_{il}$, then the probability of conversion into mild cognitive impairment (MCI) or dementia for that subject at visit j is denoted by p_{irj} and is greater than 0.5. Conversely, if $\mathrm{HR}_{irj} < 1$, then $p_{irj} < 0.5$.

The accuracy was then checked using a cross-validation process to make sure it was satisfactory and to guard against overfitting. The method consists of two steps: an inner step that chooses the optimal feature subsets for the models and an outer step that estimates the model's performance objectively (see overfitting). It was a 10-fold k-fold cross-validation technique design. The minimal-redundancy-maximum-relevance (mRMR) method was used to choose the optimal feature subsets. The inner stage involves repeatedly using the mRML algorithm, randomly partitioning the training data, and recommending feature combinations 100 times. The best three feature combinations were then chosen, and 30,000 assessments of chosen subsets for each dimension were conducted in the outer stage. Sensitivity, specificity, accuracy, and the area under the ROC curve were used to evaluate the models. On the basis of their frequent occurrences in the feature subsets and superior classification scores, the top models were finally picked. See Figure 2.5.

Disease progression models

While the Long-Term Joint Mixed Model (LTJMM) mandates that long-term trajectories must be linear, GRACE permits many monotonous curve forms without specifying any parametric families. The original values of the outcomes were converted into percentiles using a weighted empirical cumulative distribution function prior to fitting the DPM techniques to the data. This makes sure that the results were orientated to be growing and on a common scale, which are the indicators of

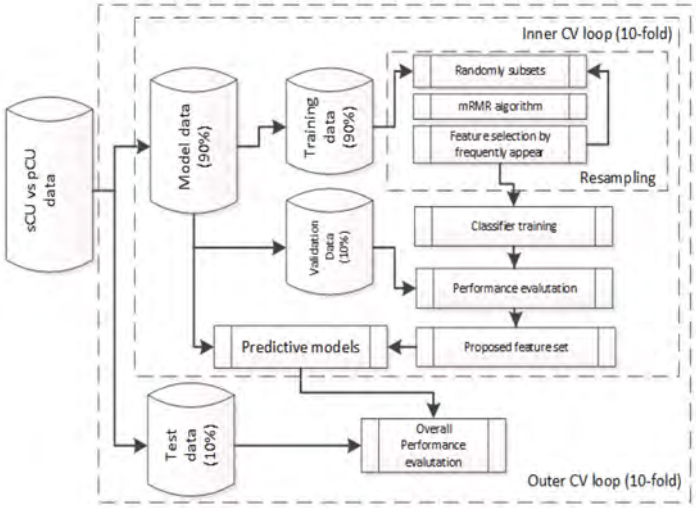

Figure 2.5: Cross-validation procedure
[50]

a disease's progression from normal to abnormal. To create a uniform scale, the percentile scale was chosen. The weighted empirical cumulative distribution function was used to generate the percentiles. The old scale was then changed using the expected values on the new scale.

GRACE method

A new component is added the to mixed effects modeling represented by g_k which is a continuously differentiable monotone function and δ_i which is the subject-specific time shift with mean zero and variance σ_δ^2. The short-term observation time is represented by $t_{ijk}^c = t_{ijk} - (t_{iend}/2)$ and the response variable is given by:

$$y_{ijk} = g_k\left(t_{ijk}^c + \delta_i\right) + x'_{t_{ijk}^c}\beta_k + \alpha_{0ik} + \alpha_{1ik}t_{ijk}^c + e_{ijk}$$

where $t_{iend} = max(q_{ik})$ and k represents the long-term progression time which is computed from $t_{ijk}^c + \delta_i$. A self-modeling regression model is applied with linear subject-level effects and long-term features with nonparametric monotone smoothing. The goal of the algorithm is to estimate both the time shift parameters and the short-term and long-term curves.

LTJMM method

The parameter γ_k corresponds to the outcome-specific slope with respect to the shifted or long-term time $t_{ijk}^c + \delta_i$. The time shift δ_i quantifies the progression of the i-th individual relative to the population and is assumed to follow a normal distribution with mean 0 and variance σ_δ^2 ($\delta_i \sim N(0, \sigma_\delta^2)$). The distribution assumptions for random effects are multivariate Gaussian, such that $\alpha_i \sim N(0, \Sigma_\alpha)$.

$$y_{ijk} = \gamma_k\left(t_{ijk}^c + \delta_i\right) + x'_{t_{ijk}^c}\beta_k + \alpha_{0ik} + \alpha_{1ik}t_{ijk}^c + e_{ijk}$$

Time zero

The study assumed that the time shift, δ_i, follows a normal distribution with mean 0 and variance σ_δ^2, and the center years of visits are represented by t_{ijk}^c. The time shift measures the relative progression of the disease in the training population, taking into consideration the variability of biomarkers. However, the time of onset may be biased due to the unequal representation of MCI subjects and CU subjects in the population. For a designated year of cognitive decline, or time zero, t_{onset}, the marker trajectories of sCU subjects should be located to the left of t_{onset}, with evolution in negative long-term times. Conversely, the marker trajectories of pCU subjects should cross t_{onset} and move towards positive values. The proposed temporal ordering allows for the measurement of sensitivity and specificity, with specificity being defined as the percentage of sCU subjects whose last visits ($t_{iend}^c + \delta_i$ where $i_{end} = \max(q_{ik})$) had negative times compared to the total number of sCU subjects.

$$SPE = \frac{\left\{ i \mid (i \in sCU) \cap \left(\left(t_{i_{\text{end}}^c}^c + \delta_i \right) < t_{\text{onset}} \right) \right\}}{sCU}$$

Sensitivity was established using two measures: (a) the proportion of pCU subjects who did not show cognitive decline at baseline and whose first visit ($t_{ci1} + \delta_i$) was negative. The proportion was calculated with respect to the total number of pCU subjects ($i1$).

$$SEN_1 = \frac{\left\{ i \mid (i \in pCU) \cap \left(\left(t_{i_1}^c + \delta_i \right) < t_{\text{onset}} \right) \right\}}{pCU}$$

and (b) the ratio of pCU subjects whose last visits have positive times with respect to the total number of pCU subjects:

$$SEN_2 = \frac{\left\{ i \mid (i \in pCU) \cap \left(\left(t_{i_{\text{end}}}^c + \delta_i \right) > t_{\text{onset}} \right) \right\}}{pCU}$$

Thus, a time zero was estimated utilising the maximisation of the three previous classification measures.

2.4.3 Results

The objective of this study was to develop cognitive decline prediction models using neuropsychological tests, cerebrospinal fluid (CSF) biomarkers, and MRI measurements of brain anatomy. Five MRI brain structure images, thirteen NMs, and five CSF biomarkers were used in the investigation. Due to the amount of assessments they underwent and the accuracy of their predictions, the best predictive models were chosen. The results show that the categorization scores of the two populations were comparable. It was advised that the ADAS11, FAQ, and EcogSPTotal scores as well as the normalized hippocampus volume be incorporated into the prediction models. Additionally, the existence of pTAU or the ratio of pTAU/Aβ or TAU/Aβ were taken into account when CSF markers were available. See table 2.5. The results demonstrated that scores increased with time, particularly from the baseline (0m) to the 12-month point, but there was no discernible change when age, gender, and years of schooling were taken into account. The proposed prediction models were also tested using the previously observed DPM algorithms GRACE and LTJMM,

Marker	coefficient	p-value
NHV	-0.33 (−0.34 −0.33)	0.019 (0.015 0.023)
NEV	-0.10 (−0.11 −0.09)	0.470 (0.377 0.563)
ADAS11	0.41 (0.40 0.41)	0.000 (0.000 0.000)
FAQ	0.29 (0.28 0.30)	0.005 (0.004 0.006)
EcogPtTotal	0.13 (0.12 0.13)	0.189 (0.151 0.226)
EcogSPTotal	0.07 (0.06 0.08)	0.532 (0.427 0.638)
Age	0.25 (0.24 0.25)	0.067 (0.054 0.081)
NHV	-0.30 (−0.31 −0.29)	0.068 (0.054 0.081)
ADAS11	0.48 (0.47 0.49)	0.001 (0.001 0.001)
FAQ	0.47 (0.46 0.48)	0.000 (0.000 0.000)
EcogSPTotal	0.25 (0.26 0.24)	0.111 (0.089 0.133)
pTAU/Aβ	0.25 (0.24 0.26)	0.012 (0.009 0.014)
Age	0.28 (0.28 0.29)	0.082 (0.066 0.099)

Table 2.5: Markers coefficients
[49]

Data	SEN (%)	SPE (%)	ACC (%)	AUC	Frequency	Optimal feature subsets
$NM + MRI_{bl}$	63.7(62.9 64.6)	64.3(63.8 64.7)	64.1(63.8 64.5)	0.688(0.683 0.693)	742–1512	*H, E, A11, F, EP, ES* *M, E, A11, F, ES* *E, F, EP*
$NM + MRI_{m12}$	77.9(76.9 78.9)	75.0(74.5 75.6)	74.7(74.2 75.1)	0.814(0.808 0.820)	2382–2821	*H, E, F, A11, EP, ES* *E, MT, A11, F, ES* *E, F, EP*
$NM + MRI_{m24}$	74.8(73.7 76.0)	77.3(76.7 77.9)	75.4(74.9 75.9)	0.822(0.814 0.829)	2170–2495	*H, E, F, A11, EP, ES* *H, A11, F, EP* *H, E, A11, F, EP*
$NM + MRI + CSF_{bl}$	64.4(63.6 65.3)	70.9(70.5 71.3)	69.2(68.9 69.6)	0.732(0.727 0.737)	2176–2229	*H, A11, F, M, ES, PT* *A11, F, ES, PTAb* *A11, F, ES, PT*
$NM + MRI + CSF_{m12}$	82.4(81.0 83.7)	71.1(70.2 71.9)	72.3(71.5 73.0)	0.833(0.823 0.842)	2176–2229	*H, A11, ES, TAb* *A11, ES, PTAb* *H, A11, EP, PTAb*
$NM + MRI + CSF_{m24}$	84.3(82.4 86.2)	74.7(73.2 76.2)	75.4(74.2 76.6)	0.854(0.840 0.867)	2176–2229	*H, A11, F, ES, PTAb* *A11, F, ES, TAb* *E, A11, EP, PTAb*

Table 2.6: Scores for predictions
[48]

and the outcomes agreed with the doctor's diagnosis. Based on the categorization of individuals and clinical groups, as well as the assessment of conversion times for patients exhibiting prodromal cognitive deterioration, the effectiveness of the DPM algorithms was assessed. While the anticipated time zero for the second group (NM + MRI + CSF) was 2.6 years with both GRACE and LTJMM, it was 1.4 years with GRACE and 3.6 years with LTJMM for the first cohort (NM + MRI).

2.4.4 Discussion

The researchers suggest a two-stage, data-driven framework for modeling Alzheimer's disease disease progression and categorical diagnostic prediction. The approach was used to assess whether the conditions of a population recruited by the Alzheimer's Disease Neuroimaging Initiative (ADNI) who were initially identified as cognitively unimpaired had later progressed to moderate cognitive impairment or dementia. In order to further explore the relevance of amyloid pathology, the method was further applied to a subset of the original data containing people having amyloid and tau information made accessible by CSF biomarkers.

By merging a limited fraction of MRI-based data, CSF markers, and conventional cognitive tests, the researchers created predicted models of CU-to-MCI/Dementia development. The developed models make use of longitudinal data. The minimum redundancy maximum relevance (mRMR) approach was used to preselect feature subsets of various dimensions, and a resampling method was then used to identify

the feature subsets that appeared most frequently for each dimension. The proposed feature subsets were assessed for cross-validated classification accuracy, and the correlation matrices between the random intercepts and random slopes from the chosen markers show that these measures support the other approaches and also provide additional insight into the issue. Numerous biomarker combination techniques that are utilized for clinical categorization and subject selection for clinical trials have been the topic of prior investigations. The pTAU/A*beta* ratio and other CSF indicators somewhat increased classification accuracy. It is important to confirm the estimated dates of progression because the time zero was determined by the clinical categorization of patients or the concept of preclinical AD. The pTAU/A*beta* ratio closely matches PET categorization and estimated clinical progression, and values over 0.028 denote successful outcomes for A. The study offers a solid basis for creating verifiable and precise prediction models that can help identify patients who experience cognitive decline more quickly and facilitate the selection of patients for clinical trials.

2.5 Longitudinal survival analysis and two-group comparison for predicting the progression of mild cognitive impairment to Alzheimer's disease (Platero and Tobar. 2020) [51]

2.5.1 Introduction and Materials

A crucial goal in the treatment of Alzheimer's disease is mild cognitive impairment (MCI), which is a stage between healthy aging and dementia. But MCI is a diverse condition with a range of clinical consequences. To correctly identify MCI patients who may develop Alzheimer's disease, it is crucial to comprehend Alzheimer's disease and how it progresses. Clinical choices on treatment plans and the early identification of patients at risk can be aided by accurate diagnostic prediction with high sensitivity and specificity. The course of Alzheimer's disease has been examined using a variety of disease indicators, such as blood tests, neuropsychological evaluations, and neuroimaging biomarkers. While some studies have indicated that neuroimaging, and more especially structural MRI-based indicators, support an earlier and more accurate MCI-to-Alzheimer's diagnosis, some have showed that baseline cognitive measures have high power in predicting MCI progression to Alzheimer's disease. Combining these many methods may enhance the effectiveness of early diagnosis.

Individual patterns of change can be seen in longitudinal research, which provides pertinent data that can help narrow the differential diagnosis. For the purpose of overcoming the limitations of group comparison based on categorizing MCI participants into converters and non-converters, survival models take into account a distinct clinical group that accounts for conversion timeframes and finite follow-up or censoring. Cox proportional hazards regressions were used to create predictive models for the transition from MCI to Alzheimer's. The validity of these models depends on the exploratory variables, which may not hold true for indicators that may distinguish between individuals with stable MCI and those who are experiencing progressive MCI.

In patients with varied numbers of clinic visits, the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset was used to assess prediction models of MCI-to-AD conversion. The dataset includes three years' worth of longitudinal brain T1-weighted MRI data from MCI patients who were monitored. The photos were preprocessed using the N3 technique and B1 bias field correction. Patients were divided into two groups - those with stable MCI (sMCI) and those who converted to probable AD (pMCI) within the three-year follow-up period.

2.5.2 Methods

The goal of the study was to use a combination of brain imaging, cognitive assessment, and machine learning approaches to predict the development of moderate cognitive impairment (MCI) to Alzheimer's disease (AD). A three-stage methodology, consisting of feature extraction, feature selection, and classification, was employed in the investigation. The researchers collected a collection of data from MRI scans and cognitive tests, narrowed them down using an algorithm, and then used two longitudinal classification algorithms to assess various feature subsets. Specific region-specific brain volumes and cognitive test results were among the features that were chosen. Mixed effects modeling was employed by the researchers to analyse longitudinal data and find the best feature subsets. Using a mix of brain imaging, cognitive testing, and machine learning approaches, the study offers a thorough approach to predicting the development of MCI to AD.

Longitudinal Classification

Based on the baseline and time-varying factors, the Cox proportional hazards model is used to calculate the risk of clinical events (such the conversion from MCI to AD).

$$h(t, X) = h_0(t) \exp\left(\sum_{k=1}^{p} \alpha_k \cdot X_k\right)$$

where h_0(t) is the baseline hazard function and α= (α_1, α_2, ..., α_p) is a vector of regression coefficients.

Longitudinal mixed-effects (LME) models are used to model the longitudinal trajectories of the measurements, which allow estimates of these measurements to be known at any time. The influence of sociodemographic characteristics, such as age, sex, education, and the APOE genotype, is also considered in the LME modeling. The best predictive results were obtained when constructing LME models that added age, sex, and years of education as covariates. The hazard ratio (HR) quantifies the differential risk of a subject characterized by X_S in relation to a reference subject characterized by X_R :

$$\mathrm{HR}\left(X_S, X_R\right) = \frac{h\left(t, X_S\right)}{h\left(t, X_R\right)} = \exp\left(\sum_{k=1}^{p} \alpha_k \left(X_{S,k} - X_{R,k}\right)\right)$$

If HR is greater than 1, it means that the individual characterised by X has a higher risk of disease conversion compared to the reference individual X_R, as indicated by the values of k in equation (1). On the other hand, if HR is less than 1, the risk of conversion is lower. The primary assumption underlying the Cox model is that the ratio of the hazard functions for the two groups remains constant over time.

$$L = \prod_{m=1}^{M} L_m(\alpha) \quad L_m(\alpha) = \frac{h\left(t_m, X_m\right)}{\sum_{r \in R_m} h\left(t_m, X_r\right)}$$

The method of maximizing partial likelihood is utilized to compute the α coefficients (Cox, 1975). The time conversion from MCI to AD is denoted by m and is part of the set 1, 2, ..., M, where M represents the number of converts. The partial likelihood is the product of M terms, for each calculated between the hazard functions of the subjects whose diseases converted at the mth event of time (tm) in relation to all subjects whose diseases haven't converted. The risk set at tm is known as Rm, and it indexes the subjects who remain as MCI patients at tm. The unknown α parameters of the model are then maximized with respect to the partial likelihood function. A set of equations is derived from this process, which enables the estimation of α parameters using a numerical optimization method.

Time is not a factor for variables like gender or years of schooling. While time is proportionate to age. A longitudinal study's majority of biomarkers, however, are neither stable nor proportionate to time. For both independent and dependent variables throughout time, the Cox model may be expanded:

$$L = \prod_{m=1}^{M} L_m(\alpha) \quad L_m(\alpha) = \frac{h\left(t_m, X_m\right)}{\sum_{r \in R_m} h\left(t_m, X_r\right)}$$

where the second term in the exponential includes the effects of p_2 timevarying variables $\mathrm{Y(t)} = (Y1(t), Y2(t), ..., Y_{p2}(\mathrm{t}))$ with associated coefficients $\delta = (\delta 1, \delta 2, ..., \delta_{p2})$ The time-varying variables and their associated coefficients are included in the exponential equation of the model, and the model's parameters are estimated by maximizing partial likelihood. When analyzing the extended Cox model parameters, the values of markers at each conversion time for all individuals in the risk set are required. However, measurements may be obtained at random intervals and the acquisition and conversion times may not line up in longitudinal studies. LME models are used to model longitudinal data and estimate values at any time in order to get around these challenges.

LME models are extensions of linear models that take into account the causes of variation within and between subjects and allow for both fixed and random effects (Bernal-Rusiel et al., 2013b). The LME model is written as follows:

$$Y_i = Z_i\beta + W_i b_i + e_i,$$

where Yi is the vector of a feature for subject i's time points, Zi is the design matrix for the fixed effects (which includes elements like clinical group, age, sex, education, and scan time), and are the fixed effects coefficients, which are the same for every subject. For subject-specific random effects in addition to the fixed effects, a mixed effects model is employed, where Wi is the design matrix for the random effects, bi is a vector of the random effects, and ei is a vector of measurement errors. The bi components show the deviation of the subset of regression parameters for the ith subject from the population's values.

Age, sex, education level, and other sociodemographic details were gathered. The genotype of the Apolipoprotein E (APOE), a genetic risk factor for AD, was also recorded. Also taken into account was the relationship of time and APOE genotype

status. The greatest predictive outcomes were from LME modeling without APOE genotype as a covariate, with age, sex, and education as covariates. The clinical group was taken into account via a Boolean variable and its interaction with time for a two-group comparison.

$$\begin{aligned} y_{\mathrm{ij}} = & \left(\beta_1 + \beta_2 \cdot \mathrm{Group}_i + \beta_3 \cdot \mathrm{Age}_i + \beta_4 \cdot \text{ Education }_i + \beta_5 \cdot \mathrm{Sex}_i + b_{\mathrm{ri}}\right) \\ & + \left(\beta_6 + \beta_7 \cdot \mathrm{Group}_i + b_{\mathrm{si}}\right) t_{\mathrm{ij}} + e_{\mathrm{ij}} \end{aligned}$$

where t_{ij} is the scan time from baseline (in years), y_{ij} is the jth measure of a feature from subject I, j=1,..., n indexes the time points, n indicates the number of scans for subject I and $\beta_r = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T$ and $\beta_s = [\beta_6, \beta_7]^T$ stands for slope and intercept, respectively. If the ith subject advances to AD, the boolean variable $Group_i$ is true; otherwise, it is false. Unfortunately, the LME modeling did not take into account the impact of the clinical group in the survival analysis:

$$\begin{aligned} y_{\mathrm{ij}} = & \left(\beta_1 + \beta_2 \cdot \mathrm{Age}_i + \beta_3 \cdot \text{ Education }_i + \beta_4 \cdot \mathrm{Sex}_i + b_{\mathrm{ri}}\right) \\ & + \left(\beta_5 + b_{\mathrm{si}}\right) t_{\mathrm{ij}} + e_{\mathrm{ij}}. \end{aligned}$$

Predictive models using the two-group comparison

The random vectors b_i and e_i, which adhere to mean zero-Gaussian multivariate distributions, describe the difference between the longitudinal trajectory of each subject and the LME model. These vectors are utilized to determine the longitudinal trajectory residue, and the features are trained and classified using linear discriminant analysis (LDA). No parameter adjustments are necessary because the LDA is fed the marginal residues of the markers' longitudinal trajectories as inputs.

$$l_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(y_{\mathrm{ij}} - \left(Z_i\right)_j \beta \right),$$

where (Zi)j is the design matrix's j-row vector, which is triggered by the clinical group's boolean variable (Groupi = 1), indicating the effects of AD progression relative to stable MCI.

Predictive models using survival analysis

The models were modified to take age, education, and gender into consideration. The disease conversion time for those who progressed to AD and the censoring time for those with stable MCI were both known. Extended Cox models were created at each key time point, and hazard ratios were then calculated and turned into probabilistic terms of conversion from MCI to AD using logistic regression:

$$p\left(X_{S,v}\right) = \frac{1}{1 + \frac{1}{\mathrm{HR}_v\left(X_{S,v}, X_{R,v}\right)}},$$

Where HR_v is the hazard ratio in visit v with v = 0, 12, 24, 36. The exploratory variables for the subject and the reference in the visit v are represented by the vectors XS,v and XR,v, respectively. These vectors are made up of p1 time-independent and p2 time-varying variables, with the latter being modeled by means of LME.

Features	$A13$	ADV_T	MT_T	FAQ	IT_T	E_T	RT_{im}	IP_T	H_V
p-values (C_1)	$3.1E-18$	$1.2E-12$	$7.8E-12$	$3.6E-11$	$4.8E-9$	$9.0E-9$	$7.9E-15$	$6.0E-10$	$4.0E-11$
p-values (C_2)	$3.3E-16$	$1.1E-16$	$4.8E-14$	$3.3E-17$	$2.1E-16$	$5.8E-14$	$1.8E-6$	$2.2E-6$	$1.9E-5$
Ranking	1	3	4	5	8	10	17	18	22

Table 2.7: The most significant p-values
[52]

Feature selection and building the predictive models

The study employed layered cross-validation to avoid overfitting and biased model performance predictions. The best feature subsets were selected using an inner loop, while the model's performance was objectively estimated using an outer loop. Using a 10-fold cross-validation design, the procedure was repeated with different data divisions in both loops to enhance repeatability. Using a collection of combinations of markers with different dimensions that were assessed in the outer loop, predictive models were built for each inner loop. The chosen subsets for each dimension underwent 30,000 evaluations, with the best feature combinations in terms of classification accuracy being selected for each outer iteration. The most accurate prediction models had higher AUC values and a decent ratio of sensitivity to specificity.

2.5.3 Results

Using the suggested methods, the study processed 1330 images from 321 participants. The longitudinal image processing pipeline's consistency was verified using a quality control procedure. For each visit by each individual, cortical thickness and subcortical volumes were measured using Freesurfer's longitudinal pipeline. 40 MRI predictors of cortical and subcortical regions as well as 11 neuropsychological measures were obtained during the feature extraction stage. Each of the 51 indicators was subjected to a univariate study before a multivariate analysis was carried out to produce the suggested prediction models.

Univariate analysis of the markers

In the study, marker discrimination capacities according to clinical group and time were tested using LME modeling:

$$H_0 : C\beta = 0 \quad \text{and} \quad H_A : C\beta \neq 0$$

The contrast matrix was used to test the null hypothesis, and table 2.7's p-values for the top markers that distinguished between sMCI and pMCI are displayed. Based on the sum of the two p-values, the markers were sorted, with absent markers denoting subgroups of markers or the same marker on different sides. The left hippocampus normalized volume was the first volume measurement. The smoothed longitudinal trajectories of a few of the top biomarkers from the univariate analysis were shown in Figure 2.6.

Performance of the predictive models of MCI-to-AD progression

The experiment included both a survival analysis method and a two-group comparison technique. Using the two-group comparison method, the longitudinal trajectory

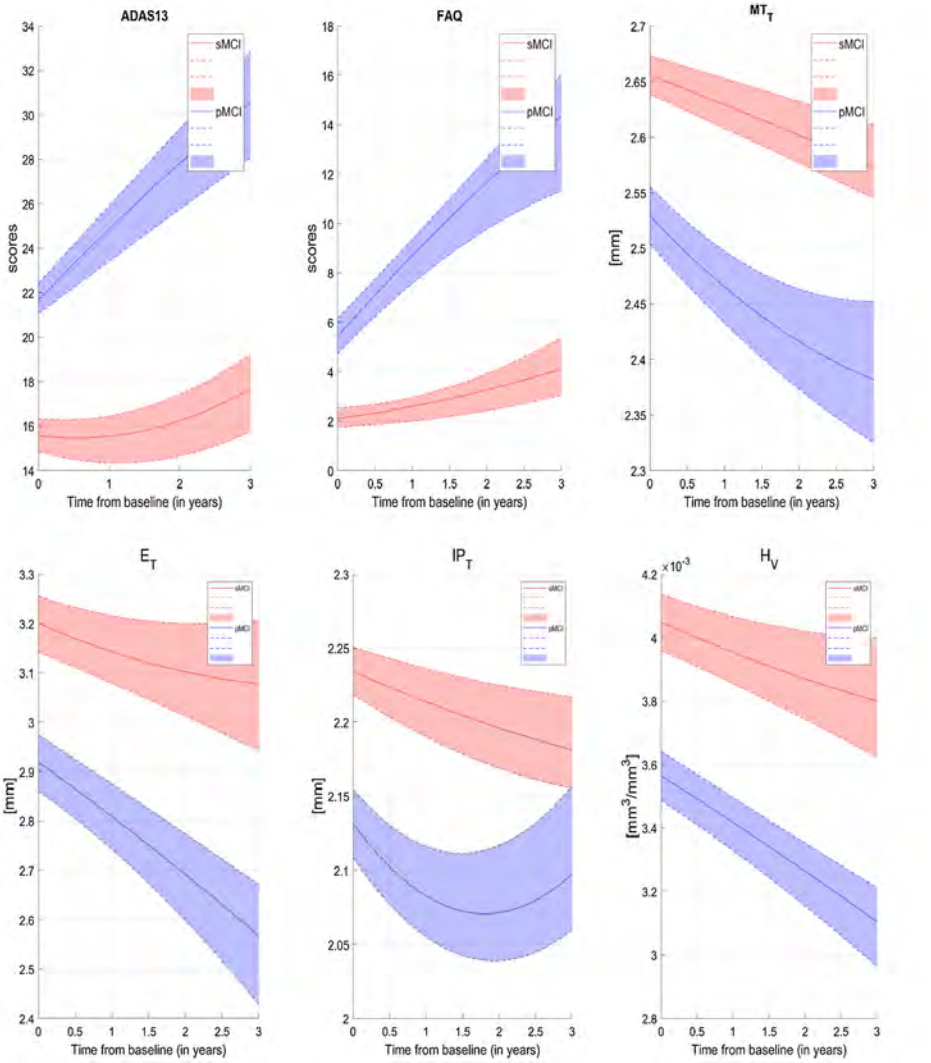

Figure 2.6: Longitudinal trajectories of biomarkers [55]

Data	AUC	ACC (%)	SEN (%)	SPE (%)	Frequency	Optimal feature subsets
MRI^{2G}_{bl}	0.769(0.763 0.775)	69.7(69.1 70.3)	74.1(73.2 74.9)	65.6(64.7 66.4)	376-981	H^L_V, P^L_V, MT^B_T H^L_V, P^L_V, IP^L_T, MT^B_T E^B_T, IP^L_T
MRI^{SA}_{bl}	0.769(0.763 0.775)	70.2(69.6 70.8)	68.8(68.0 69.7)	71.5(70.7 72.2)	756-1760	H^L_V, P^L_V, IP^L_T, MT^B_T H^L_V, P^L_V, E^B_T, IP^B_T H^L_V, P^L_V, IP^L_T, E^B_T
MRI^{2G}_{m12}	0.797(0.791 0.803)	71.3(70.7 71.9)	73.9(73.0 74.7)	69.2(68.3 70.1)	473-552	H^L_V, P^L_V, IP^L_T, MT^B_T H^L_V, P^L_V, IP^L_T, E^B_T H^L_V, P^L_V, MT^B_T
MRI^{SA}_{m12}	0.789(0.783 0.795)	70.9(70.4 71.5)	71.4(70.5 72.3)	70.7(69.9 71.6)	636-1915	H^B_V, P^L_V, MT^B_T H^B_V, P^L_V, IP^L_T H^B_V, P^L_V, IP^B_T
MRI^{2G}_{m24}	0.810(0.804 0.816)	72.4(71.8 73.0)	73.9(73.0 74.8)	71.0(70.2 71.8)	363-573	H^B_V, P^L_V, IP^L_T H^L_V, P^L_V, IP^L_T, MT^B_T H^L_V, P^L_V, MT^B_T
MRI^{SA}_{m24}	0.817(0.810 0.823)	73.7(73.0 74.3)	75.5(74.6 76.5)	71.9(70.9 72.8)	502-758	H^B_V, P^L_V, MT^B_T H^B_V, P^L_V, IP^B_T, MT^L_T, E^B_T H^B_V, P^L_V, IP^L_T
MRI^{2G}_{m36}	0.858(0.849 0.868)	74.2(73.3 75.1)	75.7(73.9 77.4)	72.9(71.7 74.2)	574-1471	H^L_V, P^L_V, IP^L_T, E^B_T, MT^B_T H^L_V, P^L_V, IP^L_T, MT^B_T H^B_V, P^L_V, IP^B_T, E^B_T
MRI^{SA}_{m36}	0.885(0.872 0.898)	78.3(76.9 79.7)	79.0(76.5 81.5)	78.1(76.2 80.0)	470-1165	H^L_V, P^L_V, IP^L_T, E^B_T, MT^B_T P^L_V, IP^B_T, E^B_T H^B_V, P^L_V, IP^L_T

Table 2.8: Scores for predicting MCI-to-AD conversion: only MRI-based biomarkers. [54]

residue of a few markers was compared between MCI persons who progressed to AD and those who did not. Based on the longitudinal trajectories of the selected markers, the survival analysis technique was used to examine the time to conversion to AD.

Both techniques used longitudinal mixed effects (LME) modeling to identify the trajectories of the selected markers for each person. The scientists found that integrating MRI-derived markers with neuropsychological measures (NMs) improved prediction accuracy compared to using only MRI-derived indicators. The results showed that during the longitudinal investigation, the extended Cox models had a better overall balance of sensitivity and specificity, and they also showed an increasing trend in sensitivity and specificity scores. While the specificity of the two-group comparison approach increased over time, its sensitivity remained stable over time at roughly 74%. The survival analysis method outperformed the two-group comparison technique after the second year.

When compared to recently published models, the suggested prediction models performed well overall and revealed combinations of markers with comparable performance. The study is significant for generating baseline ratings with a more controllable and unique feature vector than those generated by previous cross-sectional methodologies. The prediction models improved as more patient visits became available. In addition, the study included more people and visits to the ADNI database in developing the prediction models compared to past papers.

See table 2.8 and table 2.9.

Correlations between the proposed predictive models and ADAS-Cog

The ADAS-Cog, a cognitive exam that scores a person's memory, language, praxis, and orientation, is crucial for Alzheimer's disease clinical trials. The test's longitudinal score can be used to accurately estimate how long it will take MCI patients to

Data	AUC	ACC (%)	SEN (%)	SPE (%)	Frequency	Optimal feature subsets
$(\text{MRI + NM})^{2G}_{bl}$	0.855(0.850 0.861)	75.9(75.3 76.5)	85.6(84.8 86.4)	67.1(66.2 68.0)	405-1480	P^L_V, MT^B_T, A13, FAQ, RT_{lim} P^L_V, IP^B_T, A13, AQ4, FAQ, RT_{lim}
$(\text{MRI + NM})^{SA}_{bl}$	0.860(0.856 0.864)	77.7(77.3 78.2)	79.2(78.6 79.9)	75.9(75.2 76.5)	266-461	H^L_V, MT^B_T, A13, FAQ, RT_{lim} H^L_V, P^L_V, MT^B_T, A13, FAQ, RT_{lim} H^L_V, P^L_V, IP^L_T, A13, AQ4, FAQ, RT_{lim}
$(\text{MRI + NM})^{2G}_{m12}$	0.894(0.890 0.897)	80.1(79.6 80.5)	82.5(81.9 83.1)	77.9(77.2 78.5)	709-1080	P^L_V, IP^B_T, A13, FAQ, RT_{lim} P^L_V, IP^B_T, A13, FAQ, RT_{lim} P^L_V, MT^B_T, A13, FAQ, RT_{lim}
$(\text{MRI + NM})^{SA}_{m12}$	0.891(0.886 0.895)	79.3(78.7 79.9)	78.5(77.5 79.4)	80.2(79.3 81.1)	479-635	H^B_V, P^L_V, IP^B_T, A13, AQ4, FAQ, RT_{lim} P^L_V, IP^B_T, A13, FAQ, RT_{lim} P^L_V, MT^B_T, A13, AQ4, FAQ, RT_{lim}
$(\text{MRI + NM})^{2G}_{m24}$	0.908(0.905 0.911)	81.6(81.2 82.0)	77.8(77.1 78.5)	84.7(84.1 85.3)	646-1386	P^L_V, IP^B_T, A13, AQ4, FAQ, RT_{lim} P^L_V, IP^L_T, A13, FAQ, RT_{lim} P^L_V, E^B_T, IP^B_T, A13, AQ4, FAQ, RT_{lim}
$(\text{MRI + NM})^{SA}_{m24}$	0.925(0.920 0.930)	83.6(82.9 84.2)	85.2(84.3 86.2)	82.4(81.4 83.3)	525-817	P^L_V, MT^B_T, A13, FAQ, RT_{lim} P^L_V, IP^L_T, A13, FAQ, RT_{lim} P^L_V, MT^B_T, A13, FAQ, RT_{lim}
$(\text{MRI + NM})^{2G}_{m36}$	0.917(0.908 0.926)	82.6(81.6 83.6)	74.2(72.1 76.2)	87.8(86.6 89.1)	672-1162	P^L_V, IP^B_T, A13, FAQ, RT_{lim} P^L_V, MT^B_T, A13, AQ4, FAQ, RT_{lim} H^B_V, P^L_V, IP^B_T, A13, AQ4, FAQ, RT_{lim}
$(\text{MRI + NM})^{SA}_{m36}$	0.944(0.949 0.949)	85.3(84.7 86.0)	86.3(85.2 87.4)	85.6(84.8 86.5)	692-1639	P^L_V, MT^B_T, A13, FAQ, RT_{lim} P^L_V, MT^B_T, A13, FAQ, RT_{lim} H^L_V, P^L_V, MT^B_T, A13, FAQ, RT_{lim} P^L_V, MT^B_T, A13, AQ4, FAQ, RT_{lim}

Table 2.9: Scores for predicting MCI-to-AD conversion: only multisource biomarkers. [53]

progress to AD. The combination of the ADAS13 with additional indications, such as MRI-based markers and NMs, increased the accuracy of predictions for the conversion of MCI to AD. The ADAS13 is the most reliable approach for discriminating between sMCI and pMCI patients. In multisource models that combine ADAS13 with other markers, the association between ADAS13 scores and predictive models rose over time. The prediction models with the best sensitivity and specificity ratios were those that exclusively employed MRI data.

2.5.4 Discussion

The authors draw attention to past studies that focused on a single biomarker, such as shrinkage rates in specific brain areas, which does not sufficiently account for the complexity and diversity of AD. The two objectives of longitudinal studies of MCI patients are to maximize prediction accuracy and to identify a limited number of interpretable signs that aid in understanding the course of AD. Predictive models were created using longitudinal data supplied by MCI patients. The mRMR method was used to extract a large number of cortical and subcortical features from the results of neuropsychological tests, MRI data, and feature subsets of different dimensions. The most frequent feature subsets were then analyzed for each dimension using a resampling method. The best feature subsets for the final models were chosen based on how frequently suggested features were evaluated using the CV method, better AUC values, and the optimal balancing of sensitivities and specificities.

Comparing two clinical groups—stable MCI (sMCI) and progressive MCI (pMCI)—and employing models based on survival analysis were the two approaches used to build longitudinal prediction models. The longitudinal trajectories of the markers were modelled using LME in both methodologies. However, the two-group comparison strategy considered differences between converter and non-converter individuals, whereas the survival analysis only examined data up to conversion or censoring

dates. The analysis of the MMSE of the investigated MCI population over a period of 36 months revealed that the pMCI group was not homogenous with regard to conversion time, which undermined the fundamental tenet of the two-group comparison approach. The AD-prediction markers were reevaluated and statistical survival analysis approaches were used to overcome these problems. The Cox proportional hazards model and LME models were linked to examine the relationship between time-dependent markers and the timing of conversion to AD or the censure times of the samples. In both instances, a comparatively modest collection of easily comprehendible characteristics that created reliable, robust, and trustworthy prediction models were chosen. Additionally, it was easy to add the patients' age, sex, and number of years of schooling as variables in the prediction models. In order to fairly evaluate the performance of the prediction models, nested CV loops were utilized. Two nested CV loops were used in the process: an inner loop to choose the optimal feature set for the recommended models, and an outside loop to provide an impartial evaluation of model performance. The use of k-fold CV was suggested as a way to uniformly evaluate prognostic models of MCI-to-AD development. Instantaneous risk assessment is made possible by the dynamic prediction frameworks' capacity to update the predictive models whenever fresh longitudinal measurements for the target patients become available. According to the authors, the recommended models outperform models based on lone markers, and the combination of biomarkers and cognitive tests can enhance the prediction of the change from MCI to AD.

One of the study's flaws is that the outputs of the prediction models could have been altered by the choice of a certain cohort. Additionally, more samples could be employed. It is customary to use a linear function to characterize dynamic changes in structural MRI-based markers, but more complex modeling could result in predictions from predictive models that are more accurate. The probable AD diagnosis has an accuracy of 70–90% compared to pathological diagnosis, and the research's criteria for dividing MCI patients into stable and converter groups are not especially unique. The three-year follow-up period utilized in the study makes it simpler to compare the various suggestions. When predicting the onset of AD, predictive models with follow-up periods longer than three years are less reliable.

2.6 Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review (Ansart et. all 2021) [16]

2.6.1 Introduction

Machine learning algorithms have been used to identify and forecast clinical status in Alzheimer's disease, which emphasizes the importance of early diagnosis. It's crucial to distinguish between moderate cognitive impairment (MCI) sufferers who will develop Alzheimer's and those who won't. This paper covers techniques for utilizing machine learning to forecast this evolution and suggests a systematic and quantitative examination of these investigations. The review will outline the various features of the suggested approach, analyze the data to determine which features have the greatest influence, and offer suggestions for assuring the algorithm's applicability in clinical settings.

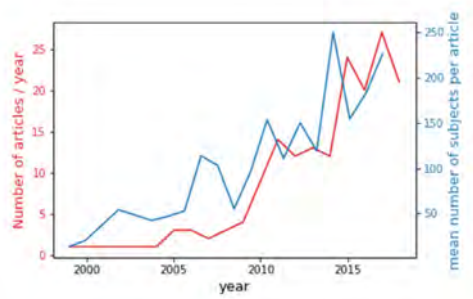

Figure 2.7: Recent trends
[18]

2.6.2 Materials and methods

The four-part query used to pick the paper on Scopus produced the identification of 330 articles, which were then processed and reduced to 172 articles by removing unrelated papers. Only articles (sMCI and pMCI) with more than 30 people in each category were thoroughly read, yielding 111 articles for analysis. One of 19 readers evaluated each of the 111 articles before one of the authors completed a final curation to ensure homogeneity. A table including the papers and reported values may be obtained on the provided website. A total of 234 trials were looked at. To prevent having a detrimental effect on the experiments, the problems mentioned in each article were taken from the table.

The text outlines a number of methodological problems that were discovered while reading studies on the classification of people with various types of mild cognitive impairment (MCI) using machine learning algorithms. These problems include the absence of a test data set, feature embedding on the entire data set, data leakage, automatic feature selection on the entire data set, and choosing the input visit for pMCI patients based on the date of their AD diagnosis. The authors further point out that additional methodological problems that did not fall under these headings were also observed, like incompatibility between various reported metrics. The authors state that only articles with no reported issues were used when analyzing the performance of the methods, but acknowledge that some issues may not have been detected or identified during reading.

2.6.3 Descriptive analysis

the number of articles published on the prediction of MCI to AD dementia has been steadily increasing since 2010, along with an increase in the number of individuals used for experiments, with 84.6% of articles using data from the ADNI study. The reported AUC is also increasing over time, which can be attributed to new algorithms and features, as well as the use of larger data sets. See figure 2.7 and figure 2.8.

The most frequently used feature is T1 MRI, which is followed by cognitive tests and socio-demographic characteristics. Less often used biomarkers in research include FDG PET, APOE, and CSF AD. Whereas neuro-psychological tests often assess many cognitive domains, studies using T1 MRI mostly concentrate on particular brain regions of interest or the entire brain. Since T1 MRI researchers frequently originate from the medical imaging community, its ubiquity is not surprising. The

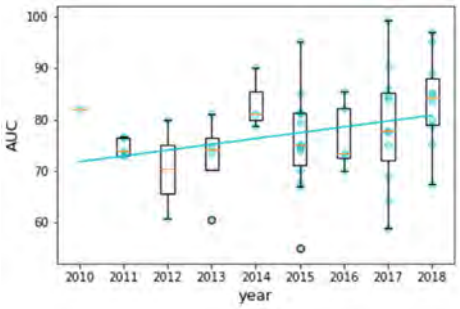

Figure 2.8: Recent trends
[17]

level of familiarity with them in the imaging community may also have an impact on the selection of cognitive characteristics. In none of the cited research was functional MRI employed.

Support Vector Machines (SVM) and logistic regressions are the most often employed algorithms in studies of Alzheimer's disease; they are utilized in 32.6% and 15.0% of instances, respectively. 63.2% of the SVM studies employ non-linear kernels, whereas 30.3% use linear kernels. Only 10% of the time do other algorithms get used. While neural networks have just been employed in the previous two years, random forests have been in use since 2014. The profession has been sluggish to adopt new algorithms even though the percentage of SVM has been declining since 2013 on average. This can be the case since the algorithm of choice has little to no effect on performance.

17.5% of Alzheimer's disease-related experiments use leave-one-individual validation, while 29.1% use the 10-fold cross-validation approach. The same subjects are used in about 7.3% of investigations, whereas 7.3% train one cohort and test another. The algorithm's performance may be impacted by the cross-validation method used, with a bigger training set and smaller test set being preferable. To compare the outcomes of various investigations, reporting variance or confidence intervals is crucial; however, this data was not gathered in the current study.

2.6.4 Performance Analysis

A linear mixed-effect model was utilized for the performance analysis in the study to assess how different method characteristics affected the AUC (area under the curve) of the prediction models. The findings demonstrated that the AUC was significantly improved by the use of EEG and MEG, domain-targeted cognitive characteristics, FDG PET, or APOE. However, neither the algorithm type nor the subject count demonstrated a substantial impact, nor did the utilization of the ADNI cohort and longitudinal data. The use of T1-ROI features, FDG PET features, and domain-targeted cognitive features had a significant effect on the AUC when just the studies carried out on the ADNI cohort were examined. The study compared the effects of using each feature individually to using them all together. The findings demonstrated that employing T1 MRI in combination with other features had a considerably better effect on the AUC than doing so with just T1 MRI. However, the correlation between cognition and FDG PET was not significant.

Cognition, medical Imaging and biomarkers

The study demonstrated that the performance of AD diagnosis methods was greatly enhanced when cognitive factors were included to T1 MRI. The performance of other modalities, including EEG, MEG, FDG PET, and APOE, also significantly improved. However, compared to using cognitive variables alone, using additional modalities in combination with cognitive variables did not significantly improve performance. Consequently, approaches that are only cognition-focused, especially those that incorporate domain-specific cognitive scores, ought to be further investigated because they are affordable and have a history of improving AD diagnosis.

The most accurate modality for predicting the development to AD, according to the researchers, is FDG PET, which is followed by cognitive characteristics and T1-ROI features. T1 MRI alone did not perform as well as other modalities, and neither amyloid PET nor CSF value significantly affected the accuracy of the predictions. Although Tau PET has not been thoroughly researched, early detection is predicted to benefit from it. Performance may also be significantly impacted by the use of EEG or MEG, but further study is required to prove this. Therefore, the authors contend that although cognitive features may be a more cost-effective option for routine clinical use, imaging methods can still offer insightful information on the progression of disease.

Longitudinal data, algorithms and other methodological characteristics

The evolution of AD can be better understood with the help of longitudinal data, which may also enhance prediction accuracy. However, the review study found that performance was unaffected significantly by the use of longitudinal data. This is in line with earlier research on AD and other degenerative disorders in general. Since that there is no distinct temporal marker of illness progression prior to diagnosis, designing longitudinal research in age-related disorders can be difficult. Moreover, patients are visited at various intervals, and not all features are acquired during each appointment, which causes missing values. Hence, compared to cross-sectional techniques, methodologies for longitudinal analysis in AD are more exploratory. It is important to remember that longitudinal data might still be beneficial for comprehending the disease process and locating potential AD biomarkers. To further understand how to include longitudinal data into AD prognostic models, more study is nonetheless required.

Table 2.10 demonstrates that there is no appreciable difference in performance depending on the algorithm used, with non-linear models having a marginally higher coefficient than linear and modified linear models. Unfortunately, none of the model coefficients are statistically significant. The authors take into account how the model selection and the application of imaging features interact (see table 2.11). The interaction between linear models and imaging characteristics is highly positive, demonstrating that adding imaging features can significantly improve performance when using a linear model, even if linear models perform noticeably worse than other models. Similar outcomes are shown for generalized linear models, but the impact on non-linear models is negligible. The authors draw the conclusion that non-linear models, which do not greatly benefit from the addition of imaging data, yield the best results when the various coefficients are combined. Both the use of the ADNI data set and its impact on data set size are examined by the authors,

Characteristic	coeff.	*p*-value	corrected *p*-value	number of exp.
intercept	78	**0**	**0**	NA
linear model	−0.47	0.79	0.94	23
generalized linear model	0.19	0.89	0.96	28
non linear model	0.83	0.55	0.79	50
T1 features	0.92	0.26	0.76	77
amyloid PET	1.3	0.35	0.79	5
FDG PET	2.6	**0.023**	0.13	24
white matter hyper-intensities	−0.58	0.49	0.79	3
EEG/MEG	3.4	$\mathbf{2.9{*}10^{-03}}$	**0.029**	5
general cognitive features	−0.14	0.91	0.96	49
domain targeted cognitive features	2.6	**0.026**	0.13	25
new or specific cognitive features	0.89	0.52	0.79	2
socio-demographic features	1.2	0.43	0.79	43
APOE	2.27	**0.049**	0.19	26
biomarkers	0.75	0.39	0.79	19
other features	0.53	0.54	0.79	12
longitudinal	0.25	0.80	0.95	13
ADNI	0.011	0.99	0.99	106
number of subjects	−0.39	0.76	0.94	NA
individual intercept	NA	0.072	0.24	NA

Table 2.10: Impact of method characteristics
[14]

Characteristic	coeff.	*p*-value	corrected *p*-value	number of exp.
intercept	78	**0**	**0**	NA
linear model	−8.1	$\mathbf{1{*}10^{-03}}$	$\mathbf{5.5{*}10^{-03}}$	23
generalized linear model	−3.7	0.12	0.25	28
non linear model	−0.13	0.96	0.96	50
imaging features	−1.6	0.42	0.52	94
cognitive features	2	**0.028**	0.073	53
socio-demographic features and APOE	2.4	**0.012**	**0.04**	49
biomarkers	0.92	0.28	0.45	19
other features	0.87	0.31	0.45	12
longitudinal	0.35	0.72	0.82	13
ADNI	−1.42	0.26	0.45	106
number of subjects	−0.066	0.96	0.96	NA
interaction: linear model and imaging features	7.85	$\mathbf{5.8{*}10^{-04}}$	$\mathbf{4.6{*}10^{-03}}$	19
interaction: generalized linear model and imaging features	4.41	**0.036**	0.083	21
interaction: non linear model and imaging features	2.13	0.41	0.52	38
individual intercept	2.27	$\mathbf{8.6{*}10^{-03}}$	**0.034**	NA

Table 2.11: Impact of method characteristics
[15]

who find that neither element significantly affects performance. Nevertheless, the outcomes differ slightly depending on whether all experiments or just the ADNI experiment are used. Despite using a hierarchical grouping of the variables to boost statistical power, the authors highlight that few p-values and corrected p-values are significant, indicating that the reported performance measures' variation is excessive in comparison to the effect sizes.

The performance of the classification models was not significantly affected by the number of subjects in the data set (coefficient = -0.39, p = 0.76). In Section 5.1.2, the effect of data set size was examined in more detail. The performance was unaffected by the use of the ADNI data set (coefficient = 0.011, p = 0.995), however the outcomes differed slightly depending on whether all experiments or just the ADNI experiment were used. It was challenging to determine the effect of its usage independently from other variables because only 14 of the listed research did not use the ADNI database. Even though the variables were grouped hierarchically to boost statistical power, only a small number of p-values and corrected p-values were significant. This shows that there is a large range between the stated performance metrics and the effect sizes.

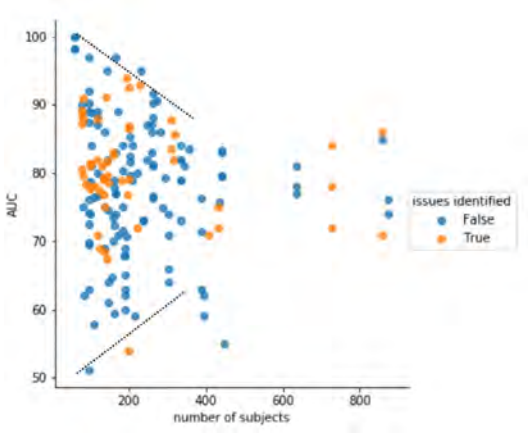

Figure 2.9: Relation between the AUC and n
[19]

2.6.5 Design of the decision support system and methodological issues

The effectiveness of machine learning models for predicting Alzheimer's disease is explored in relation to a number of concerns that have been observed in the research that have been analyzed (AD). The absence of test data sets and data leakage are the first problem, which might cause an overestimation of the model's performance. The temptation to execute feature selection throughout the entire data set is the second problem, which can also result in overfitting. An upper-limit and a lower-limit of AUC are determined by analyzing how well models perform as a function of data set size. The two lines converge to an AUC of approximately 75% as the number of subjects rises, which may indicate the genuine performance for current state-of-the-art methodologies (see figure 2.9. High-performing studies typically include more detected faults.

The use of test subject characteristics as well as the use of the diagnostic date are excluded since they may hinder the approach from being used to clinical practice. It is advised to favor approaches that forecast the precise dates of progression over those that predict the diagnosis at a specific time when it comes to the choice of time-to-prediction, which is also mentioned. Lastly, suggestions for how to assess various approaches objectively without abusing basic metrics like AUC or accuracy are discussed. According to the authors, one should use such measures carefully in order to avoid discouraging the publication of novel methodological research and obscuring the importance of understanding why some approaches are more effective than others.

2.6.6 Conclusion

A thorough and quantitative assessment of 111 articles on the automatic forecasting of the clinical status progression of people with mild cognitive impairment is presented in this article (MCI). According to the study, cognitive factors and FDG PET are more useful than other feature categories in predicting dementia progression in MCI patients. Despite the several strategies created for this imaging modality, T1 MRI alone produces noticeably lesser performance. The article offers recommendations for developing a technique that can serve as clinical decision support, em-

phasizing the value of an independent test set and pre-registering the time window. According to the review, clinical decision support systems may be more interested in forecasting future values of biomarkers or images. The study suggests looking into more focused cognitive assessments and digital assessments that patients can take at home. The essay concludes by advising prospective clinical trials to assess the efficiency of such systems.

2.7 Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification (Korolev et. all 2020) [40]

2.7.1 Introduction

The increased risk of Alzheimer's disease (AD) among those with mild cognitive impairment (MCI) is discussed in this research paper. To predict the progression of MCI to dementia, the authors created a multivariate predictive model utilizing clinical data, structural magnetic resonance imaging (MRI), and blood plasma-based proteomic data. To combine data from several sources, they employed multiple kernel learning (MKL) and a kernel-based classifier. The study focuses on probabilistic prediction, which enables categorization of MCI patients into various risk groups and defers judgment regarding the future state of dementia for instances that are unclear. By taking into account the confidence of the predictions, the paper intends to assess the prognostic model's efficacy, sensitivity to patient heterogeneity, and improvement in model performance.

2.7.2 Materials and methods

The Alzheimer's Disease Neuroimaging Initiative (ADNI) database was used in the analysis. The ADNI is an observational study that was started in 2003 to see if neuroimaging, fluid and genetic biomarkers, and cognitive tests could be used to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The study made use of baseline visit information gathered from MCI individuals recruited for ADNI-1. The website for the ADNI lists its qualifying requirements, and participants with MCI satisfied the Petersen (Mayo Clinic) diagnostic standards for amnestic MCI. There were 259 MCI subjects in the study; 139 of them were progressors (P-MCI), meaning they developed AD-type dementia within 36 months of enrolling in the trial, and 120 were non-progressors (N-MCI), indicating they did not develop dementia within 36 months of enrolling in the study. The participants in the study completed a thorough clinical examination, cognitive and functional tests, and a structural brain MRI scan. Blood was also donated by the participants for proteomic and genotyping studies on apolipoprotein E (APOE). Thereafter, subjects were given longitudinal follow-up at predetermined intervals (6, 12, 18, 24, 36 months). At each follow-up appointment, each MCI subject's clinical state was reevaluated and updated to reflect one of many results (NC, MCI, AD, or other).

The classification analyses of the study took into account a total of 186 clinical characteristics (features) as potential predictors of MCI-to-dementia progression.

Characteristic	N-MCI (n = 120)	P-MCI (n = 139)	p-value
Age, years	74.8 ± 7.6	74.8 ± 7.1	>0.5[a]
Education, years	15.7 ± 2.9	15.6 ± 2.9	>0.5[a]
Sex, % female	28.3	38.1	0.097[b]
APOE ε4 carriers, %	41.7	68.2	<0.001[b]
MMSE score	27.6 ± 1.7	26.7 ± 1.7	<0.001[a]

Table 2.12: Subject characteristics at baseline
[41]

Risk factors and evaluations/markers were the two categories of clinical characteristics. Age, sex, education, APOE genotype, familial dementia history, risk factors for cerebrovascular illness, body mass index, and a history of psychiatric problems, alcohol misuse, head trauma, and sleep apnea were risk factors. The evaluations/markers included overall and breakdown results from several clinical, functional, and cognitive tests. The study analyzed the data using a variety of statistical and machine learning techniques to find predictors of the development from MCI to dementia. The study's findings demonstrated that the greatest predictors of the development from MCI to dementia were clinical and proteomic characteristics combined. The best predictors were specifically shown to be a mix of 19 clinical factors and 9 proteomic characteristics. See table 2.12. For the purpose of feature selection and classification analysis, the researchers employed MATLAB R2010b. To find a subset of informative features, they transformed the feature data and employed a combined filter-wrapper technique. They then developed dementia prognostic models using the probabilistic multiple kernel learning (pMKL) classification approach. Both single-kernel mode and multiple-kernel mode can be utilized with pMKL, which is analogous to the support vector machine (SVM). The Generalized Linear Model (GLM) regression framework serves as the foundation for the pMKL classifier, which makes probabilistic predictions. The multinomial probit likelihood provided by the formula below is used as the basis for the pMKL classifier, which is built on a Generalized Linear Model (GLM) regression framework:

$$P\left(Y_n = i \mid W, k_n^{\beta\Theta}\right) = E_{p(u)}\left\{\prod_{j \neq i} \Phi\left(u + (w_i - w_j)\, k_n^{\beta\Theta}\right)\right\}$$

where p(u) = N(0,1) and are the cumulative distribution function and E is the expectation with regard to the standard normal distribution. This function, given the feature data (in the form of a kernel matrix kb Y) and regression coefficients W, calculates the probability P that example n belongs to class/outcome I (as opposed to class j). The weights n that training instances used to build the model vote for a given class/outcome are reflected in the regression coefficients.

The study used a variety of data sources alone and in combination to develop and evaluate nine predictive models for categorizing specific patients into N-MCI or P-MCI groups. Models with many kernels and sources were also looked at. The accuracy of the top-performing model was then examined under various patient heterogeneity conditions, as well as the association between predicted probabilities and time to progression for P-MCI patients. Cross-validated metrics like sensitivity, specificity, balanced accuracy rate, calibration, and AUC were used to assess the performance of the model. To prevent model overfitting, a nested stratified cross-validation approach was adopted, and the models were cross-validated on all 259 subjects. For improved reproducibility, the process was done ten times. Statistical

Model (#)	V-BAR (%)	T-BAR (%)	Sn (%)	Sp (%)	AUC-ROC	CCC	$D_{OPTIMAL}$ / Total
Single Source							
CRF (1)	62.0 ± 1.4	61.8 ± 7.7	65.3 ± 12.7	58.3 ± 11.7	0.61 ± 0.12	#	1 ± 0 / 16
CAM (2)	77.9 ± 1.4	76.1 ± 7.2	76.9 ± 9.5	75.3 ± 11.2	0.83 ± 0.07	0.92 ± 0.03	15 ± 10 / 170
MRI (3)	71.4 ± 1.6	69.1 ± 8.5	68.5 ± 11.8	69.6 ± 12.4	0.76 ± 0.09	0.91 ± 0.03	10 ± 5 / 452
PPM (4)	56.0 ± 2.7	53.2 ± 10.0	51.2 ± 12.9	55.3 ± 14.1	0.54 ± 0.11	0.10 ± 0.31	40 ± 10 / 149
Multi-Source							
CONCAT (5)	79.7 ± 1.4	80.0 ± 7.3	80.3 ± 10.6	79.8 ± 10.9	0.86 ± 0.07	0.93 ± 0.02	10 ± 3 / 787
MKL-Gaussian (6)	80.3 ± 1.3	79.9 ± 6.8	83.4 ± 9.9	76.4 ± 12.3	0.87 ± 0.07	0.95 ± 0.01	10 ± 3 / 787

Table 2.13: Cross-validated performance estimates
[39]

tests were run to look for any notable variations in model pair performance.

2.7.3 Results

The efficacy of various machine learning models in predicting the progression from moderate cognitive impairment (MCI) to dementia is discussed by the authors in this findings section. They assessed the performance of nine models, including five multi-source models and four single-source models (CRF, CAM, MRI, and PPM) (CONCAT, MKL-Linear, MKL-RBF, MKL-Polynomial, MKL-Gaussian). The models' validation and test set accuracies were within 3% of one another, showing little overfitting and efficient cross-validation techniques. All of the single-source models performed better than chance-level accuracy, with CAM doing the best. All single-source models were surpassed by the single-kernel, multi-source model CONCAT, and model 6 (MKL-Gaussian) outperformed CONCAT in calibration while retaining a comparable level of accuracy. See table 2.13.

The top 10 predictors of MCI to dementia development for each model were also determined by the authors. The number of APOE $\epsilon 4$ alleles was the single-source model's most frequently chosen feature, whereas the results of three evaluations were the model's most frequently chosen feature (ADAS-Cog, FAQ, RAVLT). Measures of volume and cortical thickness in the temporoparietal brain areas were typically chosen by the MRI model. The PPM model typically chose proteins linked to lipid metabolism, immunological response, and inflammatory processes. Only CAM and MRI variables were consistently chosen as predictors in the multi-source models.

The baseline predictors of MCI-to-dementia progression between the N-MCI and P-MCI groups were then compared in a confirmatory study by the authors. According to higher scores on the ADAS-Cog and FAQ, P-MCI patients had more cognitive and functional impairment at baseline than N-MCI respondents. They also displayed baseline evidence of atrophy in the temporoparietal brain areas and had more severe verbal memory impairment, as shown by lower RAVLT scores. Overall, the authors showed the value of multi-source models for predicting progression and found a number of predictors of MCI-to-dementia progression.

2.7.4 Discussion

Predictive utility and predictors

The most reliable predictors were discovered to be cognitive and functional tests, followed by MRI measurements and clinical risk factors. Plasma proteomic information was rarely chosen as a predictor since it had the lowest accuracy. The best prediction accuracy was produced by multi-source models that included CF

Study	Time (months)	Markers	AUC-ROC	Acc (%)	Sn (%)	Sp (%)
Present study	36	CF, MRI	0.87	79.9	83.4	76.4
Cui et al. (2011)	24	CF, MRI, CSF	0.80	67.1	96.4	48.3
Gomar et al. (2011)	24	CF, MRI	0.80	71.9	56	82
Hinrichs et al. (2011)	36	MRI, PET	0.74	—	—	—
Westman et al. (2012)	36	MRI, CSF	0.76	68.5	74.1	63.0
Ye et al. (2012)	48	CF, MRI, APOE	0.86	—	—	—
Zhang and Shen (2012)	24	MRI, PET, CSF	0.80	73.9	68.6	73.6
Wee et al. (2013)	36	MRI	0.84	75.1	63.5	84.4
Young et al. (2013)	36	MRI, PET, APOE	0.80	74.1	78.7	65.6

Table 2.14: Comparison of models for predicting MCI-to-AD progression [38]

assessment scores and morphometric MRI measurements. The most accurate model was able to more accurately identify MCI patients who had AD dementia within 18 months of their baseline than those who did not. The results imply that certain biomarkers have distinctive temporal trajectories and may be most sensitive to AD-related alterations at particular points in time. The transition from mild cognitive impairment (MCI) to dementia over a three-year period was predicted most accurately by analyzing data from a variety of sources. With a 76.1% prediction accuracy rate, cognitive and functional (CF) evaluations were determined to be the most reliable. This was more accurate than cerebrospinal fluid (CSF) biomarkers over a two-year period and structural MRI. The least reliable predictor, plasma proteomic data, had a 53.2% predicted accuracy rate, which was barely better than chance. Additionally, the study discovered that employing multi-source models like CONCAT and MKL-Gaussian increased predicted accuracy to about 80%. Our models consistently included CF evaluations and morphometric MRI measurements as predictors, showing that both data sources offer complementary knowledge on the progression of MCI to dementia. Contrarily, plasma proteomic measurements and clinical risk indicators were not frequently chosen as predictors, indicating that the information offered by these data sources about progression is redundant or limited. The baseline scores on cognitive and functional examinations, as well as morphometric data for three different brain regions, were the predictors of MCI-to-dementia progression that were found in the multi-source models (left hippocampus, middle temporal gyrus, and inferior parietal cortex). In addition to memory function scores, cognitive assessment results were chosen as predictors, which implies that cognitive impairment in more than simply memory function is a sign of dementia. The choice of functional status scores as predictors suggests that in patients with MCI, a mild but reliable impairment in functional status occurs before the onset of overt dementia. The study also discovered that the preference for the left hemisphere in the use of morphometric MRI parameters as predictors was in line with the evidence that suggests Alzheimer's disease-related atrophy happens there more quickly.

Comparison with models in literature

Table 2.14 demonstrates that when compared to recently published models, our top prediction model (AUC = 0.87, accuracy = 79.9%) performed remarkably well. We restrict this comparison to studies that employed baseline data from the ADNI dataset to forecast MCI-to-AD development within a 24-48 month follow-up period in order to make it more compatible with the current analysis. With a sensitivi-

ty/specificity differential of only 7%, our model's prediction accuracy was not only good but also fairly balanced, which contrasts favorably with recent research where this differential was as high as 48%.

Probabilistic Classification of MCI: Advantages and Applications

One of the study's distinguishing features is the use of a probabilistic method, which enables the model to offer accurate and calibrated forecasts of the probability of advancement as well as details on the time to progression for individuals with prodromal MCI (P-MCI). Clinicians may categorize MCI patients based on their risk of progression thanks to calibration analysis, which showed that the model's probabilistic predictions accurately represent the real risk of advancement. Although not having been specifically trained for this goal, the study also discovered that the model's probabilistic predictions might contain some information on the P-MCI patients' time to progression. As a result, doctors may be able to stage MCI patients along the MCI-disease Alzheimer's (AD) continuum by adapting the model to explicitly forecast time to progression.

Significantly, the study demonstrated how the model's probabilistic outputs might be used as a gauge of prediction confidence to raise its accuracy. Without taking into account information about prediction confidence, the model performed in non-probabilistic mode and attained an accuracy of 79.9%, with 83.4% sensitivity and 76.4% specificity. According to the study, it is possible to divide MCI patients into high-risk and low-risk groups using the probabilistic prognostic model that was created, allowing for the enlargement of patient samples in clinical trials. The necessary sample size to determine the impact of a possible treatment could decrease by up to 57% as a result of this. The algorithm may potentially be used to more precisely pinpoint MCI individuals at high risk for early disease-modifying drug treatment. Clinicians may decide to request additional biomarker tests in situations when the model is unable to make a strong prediction. By utilizing this prognostic model, more costly, invasive, or uncommon tests, like PET-based amyloid imaging, could be employed with greater caution, providing both patients and the healthcare industry with major advantages.

Limitations and future directions

The dependence on clinical diagnoses as the "ground truth" for AD, which may inject uncertainty and noise into the model construction process, is one issue discussed in the paper. Models created to forecast the progression from moderate cognitive impairment (MCI) to clinically-diagnosed AD can only be as accurate as the clinical diagnosis itself because the clinical diagnosis of probable AD has an accuracy of 70–90% relative to the pathological diagnosis. Furthermore, baseline clinical assessments are frequently more predictive of progression than other types of biomarkers, which may partially be explained by the use of clinical criteria to determine when MCI-to-AD progression has place. Future studies should include data from patients with proven pathological diagnoses as well as clinical diagnoses to increase the accuracy of predictive models for AD. This might aid in lowering the unpredictability and noise brought on by relying solely on clinical diagnosis. The study's relatively brief three-year follow-up period is another drawback. Although prognostic models for long-term dementia prediction are necessary, high-risk MCI individuals can still

be identified for clinical trials by using short-term dementia prediction. The majority of MCI patients who go on to develop AD-type dementia do so within the first few years of follow-up, it's also crucial to remember.

Suggestions for future research directions are also made, such as combining patient data from both amnestic and non-amnestic MCI subtypes as well as single-domain and multiple-domain MCI subtypes to increase the clinical applicability of predictive models. In addition, the study only predicted the progression from MCI to AD; nevertheless, there are many other kinds of dementia besides AD, and many dementia cases have multiple etiologies. In light of this, the probabilistic pattern classification method used in this study might be naturally extended for use in the differential diagnosis of dementia, allowing for the creation of a multi-class classifier that would assign probabilities to various forms of dementia. Including imaging markers of brain connection, such as those based on diffusion tensor imaging and resting-state functional MRI, could also help to improve the predictive models for AD. The model's capacity to recognize MCI patients who develop AD more than 18 months after baseline may be enhanced by the addition of PET-based amyloid imaging. The study used cross-validation to assess the models' prediction performance, but the next step is to externally validate the models using a different dataset.

2.8 Estimating long-term multivariate progression from short-term data (Donohue et. all 2014) [24]

2.8.1 Introduction

Various methods for estimating smooth progression or growth curves from serial observations of individuals over a biologically common time span are discussed. For example, generalized linear or nonlinear mixed effects models can be used to describe height, weight, or pharmacokinetics over time following an event of interest. However, when studying diseases that occur over long periods of time, such as Alzheimer's disease, epidemiologic studies may lack an obvious biological event that can serve as a reference "time zero." Furthermore, short-term follow-up with few observations may necessitate much simpler subject-level characteristics.

To address these challenges, a self-modeling regression (SEMOR) model with simple, linear subject-level effects is presented, while nonparametric monotone smoothing is used to model long-term characteristics. The objective is to estimate population curves for Alzheimer's disease progression over decades using a variety of outcome indicators. The Alzheimer's Disease Neuroimaging Initiative (ADNI) has followed volunteers with cognitively normal, early mild cognitive impairment, late mild cognitive impairment, and probable mild Alzheimer's disease for up to six years, collecting data such as serial magnetic resonance imaging measurements of regional brain volumes, positron emission tomography measurements of brain function and amyloid accumulation, other biological markers, and clinical and neuropsychological assessments. However, the subset of people who develop Alzheimer's disease is small, and the data set lacks novel biomarkers of primary interest in the early stages of the disease.

To solve this restriction, long-term MMSE trajectories from the "Personnes Agées Quid" (PAQUID) research are recommended to fine-tune the findings of the algorithm applied to ADNI data and convert time to indicate time until dementia on-

set. The author expects that by doing so, he would be able to better estimate the long-term biomarker progression of Alzheimer's disease and so help in treatment development and observational research.

2.8.2 Model and Algorithm

A model for individual outcomes that emerge over time is explored, with each outcome influenced by a monotone function, Gaussian residual errors, and a subject-specific temporal shift.

$Y_{ij}(t) = g_j\left(t + \gamma_i\right) + \alpha_{0ij} + \alpha_{1ij}t + \varepsilon_{ij}(t)$

The model implies that the subject-specific time shift is an unknown "health age" that can be shifted left or right relative to the actual age due to disease manifestation at various ages. The observed covariate, t, reflects short-term observation time, while $t + \gamma_i$ denotes long-term progression time, where γ_i is the unknown subject-specific time shift.

The algorithm simplifies the highly dimensional and complex problem. Each of the unknown parameters (g_j, γ_i, and α) is estimated in turn using the existing estimations of the other parameters. This procedure is repeated until the RSS converges. The technique employs three distinct forms of partial residuals, denoted $R^g_{ij}(t); R^\alpha_{ij}(t); R^\gamma_{ij}(t)$ (Table 1). If we assume that model (1) is valid, then each of the partial residuals offers an unbiased estimate of one of the unknown parameters. Conditional expectations of partial residuals are comparable, or nearly so, to target parameters. We start the algorithm by initializing $\gamma_i = 0$ and iterating through the following steps.

1. Given γ_i, estimate the monotone functions g_i by setting $a_{0ij} = a_{1ij} = 0$ and iterating the submethod below.

(a)Calculate g_i using a monotone smooth of $R^g_{ij}(t)$:

(b) Estimate $\alpha_{Oij}, \alpha_{1ij}$ by the linear mixed model of $R^\alpha_{ij}(t)$.Repeat steps a and b until convergence of the RSS for the j th outcome: $\text{RSS}_j = \sum_{it}\left[Y_{ij}(t) - \right.$

$\left. g_j\left(t + \gamma_i\right) - \alpha_{0ij} - \alpha_{1ij}t\right]^2$.

2. With current set of g_j, set $\alpha_{0ij} = \alpha_{1ij} = \varepsilon_{ij}(t) = 0$, and approximate each γ_i with the average of $R^\gamma_{ij}(t)$ over all j and t. Repeat steps 1 and 2 until convergence of the total RRS equals $\Sigma_{ijt}\left[Y_{ij}(t) - g_j\left(t + \gamma_i\right) - \alpha_{0ij} - \alpha_{1ij}t\right]^2$.

2.8.3 Results

The findings are based on an examination of data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), a long-term study of Alzheimer's disease (AD) and moderate cognitive impairment (MCI). Figure 2.10 shows longitudinal graphs of significant variables gathered during the study, such as amyloid plaque buildup in the brain, CSF and PET measurements, MRI data, cognitive evaluations, and functional activities. The study's major goal was to create a data-driven version of the progression curves proposed by Jack and colleagues, which display the important indicators of disease development on a common vertical scale from normal to abnormal, with clinical disease stage on the horizontal axis.

Because the diagnostic groups were not evenly represented in the data, the authors used a weighted percentile transformation to convert ADNI measures to a percentile scale. The resulting scale goes from 0 (least severe observed value) to

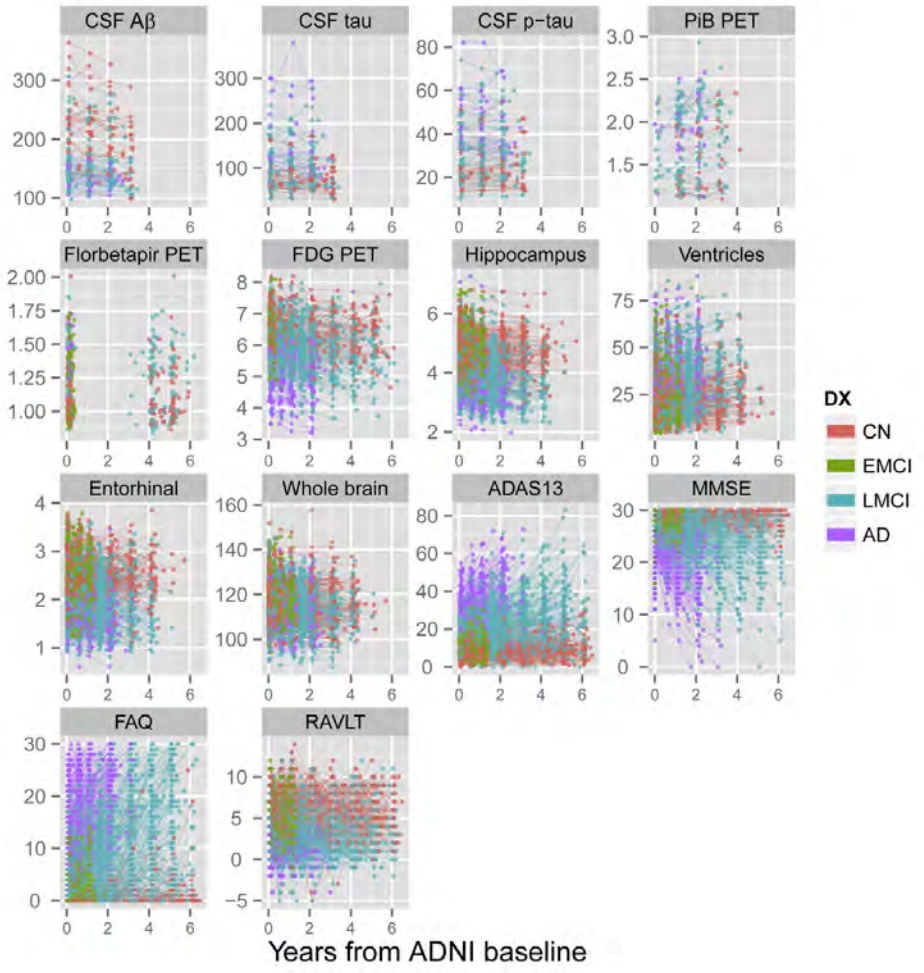

Figure 2.10: ADNI panel of biomarkers and assessments [22]

Outcome	CN, n/N*	EMCI, n/N*	LMCI, n/N*	AD, n/N*	Total
CSF tau, pg/mL	36/132		54/211	16/63	106/406
CSF p-tau, pg/mL	36/132		54/211	16/63	106/406
CSF amyloid-β, pg/mL	36/132		54/211	16/63	106/406
PiB PET, SUVR	19/49		65/141	19/34	103/224
Florbetapir PET, SUVR	270/270	290/290	252/252	98/98	910/910
FDG PET uptake	345/763	299/346	409/1263	208/434	1261/2806
Ventricles, % ICV	413/1448	298/852	548/1927	267/637	1526/4864
Hippocampus, % ICV	413/1448	298/852	548/1927	267/637	1526/4864
ADAS13	418/1883	305/825	560/2735	309/871	1592/6314
MMSE	608/2094	450/979	864/3074	474/1069	2396/7216
FAQ	416/1891	304/823	560/2781	310/907	1590/6402
RAVLT	419/1893	306/825	560/2739	312/884	1597/6341

Table 2.15: Baseline diagnosis of counts of subjects and observations [23]

100 (most severe observed value) and was generated using the empirical cumulative distribution function, weighted by the inverse proportion of observations from each diagnostic group (CN, EMCI, LMCI, and AD). Table 2.15 shows the number of patients and observations by baseline diagnostic category.

A subset of 388 ADNI subjects was employed having evidence of aberrant amyloid buildup in the brain, as assessed by CSF, PiB PET, and florbetapir PET measures, to apply their approach. B-spline smooths were fitted with five equally spaced knots and fifth-degree polynomial splines, and the algorithm was blind to diagnostic categorizations. The predicted long-term trajectories among amyloid1 ADNI participants are shown in Fig. 2.11A, with time altered so that time zero marks the period when the mean Clinical Dementia Rating Scale Sum of Boxes (CDRSB) score hits the 80th percentile. The obtained timeline can be viewed as the time required to progress to the worst CDRSB 20th percentile.

The algorithm was also applied to a sample of 570 ADNI patients who had at least one APOE4 allele. This subgroup includes numerous people who were amyloid- and were unable to be diagnosed. Time was transformed using postprocessing processes, with the PAQUID timeframe reflecting time to dementia onset and the ADNI timescale calculating subject-specific times to reach the 20th percentile of CDRSB. Overall, the study's findings imply that the algorithm can be used to predict the long-term trajectories of cognitive deterioration in people with AD or MCI. The percentile-based normalization method compares disease progression across diagnostic categories, and the algorithm is applicable to both amyloid+ and APOE 4+ subgroups.

2.8.4 Discussion

The authors compare their technique to past studies on Alzheimer's disease (AD) in the discussion section of this paper and examine the method's merits and drawbacks. They begin by addressing a recent study by Bateman and colleagues that created projected progression curves for autosomal dominant Alzheimer's disease, in which the age of onset of symptoms is expected to be near to the parent's age of onset. The authors of this article, on the other hand, are less certain about determining the age of onset in sporadic AD. SEMOR, their solution, solves this limitation by simultaneously estimating the age of onset and the progression curves using a nonparametric monotone smoothing approach. They tested their method on amyloid and APOE 4 subsets that were somewhat pathologically homogeneous. According to the authors' calculations, their iterative technique can recover reasonable predictions of long-term trajectories from short-term observations. They state that

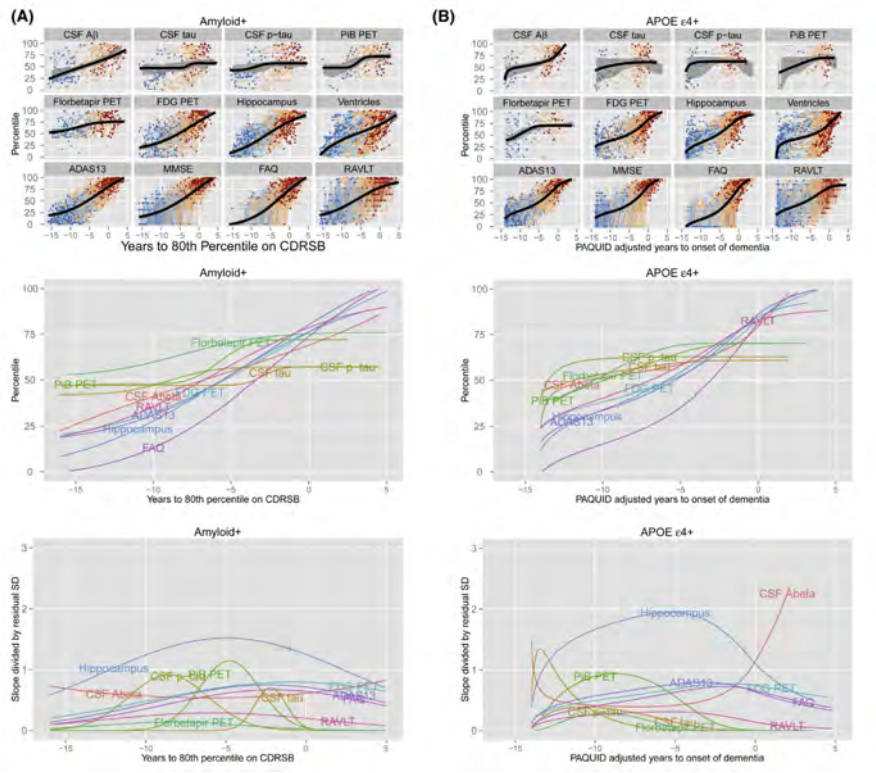

Figure 2.11: ADNI amyloid+ subjects
[21]

more simulation studies and analytical development of asymptotic convergence are needed, and that convergence of estimates of time shifts will be influenced by the number of outcomes. They also explain the approach's drawbacks, such as the fact that it does not require sigmoidal curves, but rather a very flexible class of monotone curves. They discovered that mean CSF Ab follows a linear trajectory in amyloid patients, whereas tau, p-tau, and PiB PET follow sigmoidal morphologies. Glucose metabolism, hippocampus volume, ventricular volume, learning, and cognition all follow near-linear paths. Function was the final domain to fail following a parabolic trajectory.

The authors also discuss which markers become abnormal first and show graphs of first derivatives of curves split by residual standard deviation. With the probable exception of CSF indicators, hippocampal volume appears to dominate the other metrics during the 15-year period in both analyses. The CSF markers show some places with relatively high standardized slopes, however this could be due to insufficient data and spurious acceleration around the observational boundaries. In other circumstances, the CSF measurements are rather flat, which may result in the erroneous acceleration shown by the bell shapes. Their method does not make the assumption that the mean should be between zero and one, and without this assumption, their algorithm exhibits significant pathological heterogeneity or measurement variability, even in the selected amyloid subset. They propose that food, lifestyle, education, occupation, or other factors associated to cognitive reserve, as well as genetics or family history, may explain part of the variation. They intend to test these theories further in the future by incorporating variables into the model,

although more data on the disease's early stages are required.

Finally, the comparison groups are discussed shown in their figures, pointing out that they are difficult to interpret due to the small number of subjects with known amyloid status. They recommend that future studies include data from younger cohorts and that hierarchical random effects be employed to describe within-study and within-subject association. Meta-analyses may also aid in addressing a fundamental restriction of the ADNI data, which is that the age range of ADNI participants at baseline is limited to 55 to 95 years.

2.9 Application of literature review

NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease (Jack et. all 2018)

- Definition of Alzheimer's Disease, biomarkers and cognitive stages
- AT(N) Classification

Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype (Vermunt et. all 2019)

- Estimating Alzheimer's Disease stages and comparing their view with mine.
- Comparing results of this paper with mine (Preclinical stage: 2-15 years, Prodromal stage: 3-7 years)

CSF biomarkers of Alzheimer's disease concord with amyloid-β PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts (Hannson et. all 2018)

- The research shows that in many cohorts, including BioFINDER and ADNI, CSF biomarker ratios, notably CSF total tau/Aβ(1-42) and CSF phosphorylated tau/Aβ(1-42), display significant agreement with PET categorization. This suggests that CSF biomarker tests have the potential to be trustworthy alternatives to PET in the diagnosis of Alzheimer's disease.
- In addition to being in good agreement with PET categorization, the CSF biomarker statuses generated using specified cutoffs also accurately predict more clinical deterioration in individuals with moderate cognitive impairment over a 2-year period. These findings imply that CSF biomarkers may be useful in predicting the course of the illness and may aid in the early recognition of Alzheimer's when therapies are most successful.

Categorical predictive and disease progression modeling in the early stage of Alzheimer's disease (Platero and Tobar 2022)

- The study presents a novel method for identifying subsets of markers that can precisely classify cognitively unimpaired (CU) people who are likely to develop mild cognitive impairment (MCI) or dementia using categorical predictive models based on survival analysis with longitudinal data. Using short-term clinical data, this method enables the creation of disease progression models (DPMs) that indicate long-term pathological pathways.

- The findings show that the development from CU to MCI/dementia may be accurately predicted using a limited fraction of conventional MRI-based data, cerebrospinal fluid (CSF) markers, and cognitive assessments. The developed DPMs demonstrate a significant association between anticipated conversion durations and actual clinical results using growth models and mixed effects models. The study also reveals several indicators that show substantial changes more than 20% as early as fifteen years before the start of cognitive deterioration, including temporal atrophy, clinical ratings, and the pTAU/A*beta* ratio.

Longitudinal survival analysis and two-group comparison for predicting the progression of mild cognitive impairment to Alzheimer's disease (Platero and Tobar 2020)

- The study examined two methods for building MRI and neuropsychological measurements-based longitudinal prediction models for Alzheimer's disease (AD). Both methods generated predictive models with a manageable amount of characteristics that accurately predicted AD conversion.

- In comparison to the two-group comparison technique, the survival-based prediction models showed a superior balance between sensitivity and specificity, demonstrating their efficacy in foretelling the development of AD.

Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review (Ansart et. all 2021)

- Including cognitive tests and certain imaging modalities enhanced predictive accuracy, but excluding them had no appreciable impact, according to a comprehensive analysis of research on the prediction of the development from mild cognitive impairment to Alzheimer's disease dementia.

- The review also drew attention to methodological problems, namely the lack of a test set and dubious clinical practice relevance. It stressed how crucial it is to follow ethical standards when applying machine learning to clinical decision assistance.

Estimating long-term multivariate progression from short-term data (Donohue et. all 2014)

- The GRACE model is used for the demo, to obtain the best possible combination of markers

- A novel model and estimation approach is developed to assess the timing and long-term progression of illnesses with slow progression. The technique delivers subject-specific prognostic estimations for Alzheimer's disease onset and successfully recovers disease trends.

Chapter 3

Materials

3.1 Python

Python is a popular programming language for creating many kinds of applications. Guido van Rossum developed it in the late 1980s, and it has since grown to be one of the most widely used programming languages worldwide. Python is renowned for its straightforward and uncomplicated syntax, which makes it a great language for new programmers. The enormous amount of libraries and frameworks it provides also makes it incredibly adaptable, enabling programmers to build a wide range of applications, from web apps to scientific computing to artificial intelligence.

Python is open source, which permits unrestricted use, distribution, and modification. It's accessible on a variety of operating systems, including Windows, macOS, and Linux. Python is often used in a variety of fields, including scientific computing, data science, machine learning, web development, and more. Its success is a result of a big developer community that contributes to its growth and upkeep, as well as the readability and use of the product. [9]

3.2 Jupyter Notebook

Users may create and share documents with live code, equations, visualizations, and illustrative text using the web-based interactive computing environment known as

Figure 3.1: Python logo
[8]

Figure 3.2: Jupyter Notebook logo
[4]

Jupyter Notebook. Originally created as an open-source project for the Python programming language, it has now grown to cover a wide range of additional computer languages. The three primary programming languages it supports—Julia, Python, and R—are combined to form the moniker Jupyter. Additionally, it alludes to the scientific custom of keeping notes and thoughts in a private notebook. Jupyter Notebook is made up of a server application that manages code execution and output rendering and a web application that runs on a web browser. Code and text cells may be used to build documents known as notebooks. Users may see the effects of their code as they type since code cells can be interactively executed. Equations written in LaTeX or Markdown syntax can be placed in text cells as well as styled text.

Due to its capability to combine code, data, and visualizations in a single document, Jupyter Notebook has grown in popularity as a tool for data analysis, scientific computing, and machine learning. Due to the fact that notebooks can be exported in a number of formats, including HTML, PDF, and Markdown, it enables users to quickly share their work with others. Additionally, Jupyter Notebook facilitates the development of interactive widgets that let users alter data and parameters in real-time. Overall, Jupyter Notebook is a robust and adaptable tool for scientific computing and data analysis that has grown to be a crucial component of the arsenal of the contemporary data scientist. It will continue to develop and advance in the years to come thanks to its widespread use and vibrant development community. [3]

3.3 PyCharm

Jupyter Notebook can be used directly in a browser but for this work an IDE was used, namely, PyCharm. PyCharm was created exclusively for Python programming. It is created by JetBrains and offers a complete collection of tools and features to simplify coding and increase developer productivity. Python code may be efficiently written, debugged, and tested by developers using PyCharm. It provides smart code completion, which makes it simpler to produce error-free code by suggesting pertinent code snippets, functions, and variable names as you type. The IDE also has strong navigational tools that let users easily go to certain classes, methods,

Figure 3.3: PyCharm logo
[8]

or files inside their projects. An integrated debugger in PyCharm makes it easier to find and correct coding problems. Developers may use it to set breakpoints, walk through the code, check variables, and track the course of a program as it runs. [35]

3.4 MATLAB

The high-level programming language and interactive environment known as MATLAB (short for "MATrix LABoratory") is used for numerical computing, visualization, and programming. The MathWorks company created it, and it was published in 1984. For data analysis, algorithm creation, modeling, simulation, and prototyping, MATLAB is widely used in engineering, physics, mathematics, finance, and other scientific areas. For conducting numerical operations on arrays and matrices, as well as for data visualization, image processing, signal processing, and control systems analysis, MATLAB offers a robust collection of built-in functions and tools. Additionally, it enables object-oriented programming, enabling the development of sophisticated classes and data structures.

The effective handling of massive data sets and intricate algorithms is one of MATLAB's main advantages. It is renowned for being simple to use and interactive, enabling users to play around with code and see data in real time. Additionally, a variety of toolboxes, which are sets of tools and algorithms for certain applications like statistics, optimization, and machine learning, are available in MATLAB. Both academia and industry make extensive use of MATLAB, and a sizable user and developer community actively contributes to its growth and advancement. MATLAB is a flexible tool for a variety of applications since it can be used to distribute code to standalone programs, web applications, and embedded systems. [43]

3.5 R Language

The language and environment for statistical computation and graphics is called R. The R Development Core Team presently maintains it after Ross Ihaka and Robert Gentleman first created it in 1993. R is open-source software, which entails that anybody may download, modify, and share it. Time-series analysis, clustering,

Figure 3.4: MATLAB Logo
[7]

Figure 3.5: R Logo
[10]

classification, and linear and nonlinear modeling are just a few of the statistical and graphical methods offered by R. Additionally, it includes a sizable library of user-contributed packages that expand its capability to include things like biology, finance, and social sciences.

R's adaptability and simplicity in usage for data processing and analysis are among its advantages. R has a robust collection of tools for data preparation, transformation, and cleaning, making it simple to work with a variety of data kinds and formats. In order to examine and present their data, users may construct a variety of plots, charts, and graphs using the extensive collection of capabilities it offers for data visualization. Particularly in areas like statistics, bioinformatics, and data science, R is extensively utilized in academia, government, and business. It features a sizable and active user base that aids in its growth and progress and offers users resources and assistance. [11]

3.6 ADNI

The goal of the Alzheimer's Disease Neuroimaging Initiative (ADNI) is to employ cutting-edge neuroimaging and other biomarkers to better understand Alzheimer's disease (AD). The National Institute on Aging (NIA), National Institutes of Health (NIH), private pharmaceutical firms, and nonprofit groups joined together to launch it in 2004. The ADNI study's objective is to collect information on persons with AD, moderate cognitive impairment (MCI), and normal cognition's brain imaging, genetics, clinical evaluations, and other biomarkers. The information gathered is

Figure 3.6: ADNI Logo
[2]

subsequently made freely accessible to researchers all around the world. This information is intended to increase our knowledge of AD and pave the way for the creation of brand-new therapies.

The ADNI project aims to accomplish a number of objectives, including the discovery of novel biomarkers for AD and MCI, the creation of techniques for the early detection of AD progression, and the provision of a resource for the creation and evaluation of novel AD therapies. In order to meet these objectives, ADNI gathers a variety of information from individuals, including MRI and PET scans, genetic data, cognitive tests, and clinical evaluations. We now understand a lot more about AD and MCI thanks to ADNI. It has prompted the creation of fresh biomarkers and diagnostic standards for various ailments. Numerous research studies have made use of the ADNI data, and it has even sparked the creation of brand-new AD medications and clinical trials. The ADNI project is a multi-institutional, multi-researcher collaboration that has been going on for approximately two decades. It has support from a number of institutions, including the NIA, NIH, private pharmaceutical firms, and nonprofit associations. With the aim of enhancing the lives of individuals afflicted by AD and furthering our understanding of the illness, the initiative is anticipated to last for many years to come. [1]

3.7 Margerit-3

Magerit, a computing system administered by the Centro de Supercomputación y Visualización de Madrid (CeSViMa) at Universidad Politécnica de Madrid, provides a platform for executing computationally intensive processes. This system comprises 68 Lenovo ThinkSystem SD530 nodes, each equipped with 2 Intel Xeon Gold 6230 processors featuring 20 cores per node. Moreover, Magerit is equipped with a diverse range of software and operates on the CentOS operating system. [5]

The utilization of Magerit has proven essential for the current research, as it involves time-consuming processes that would be impractical to execute on a regular computer. Thus, executing specific code on Magerit significantly reduced the overall waiting time, this was necessary to obtain the best possible combination of biomarkers for the prediction and progression model of Carlos Platero.

Figure 3.7: Magerit-3
[6]

Chapter 4

Methods

In the methods section I will focus on how I obtain my results with a primary focus on the coding aspect. As told in the materials section I will be using Python as primary coding language.

4.1 Preparing the data

There are a couple of steps that are taken each time for every figure or table or ML model and these are listed as below.

4.1.1 Importing the libraries

```
1 import numpy as np
2 import pandas as pd
3 from scipy import XXXXX
```

Listing 4.1: Importing the libraries in Python

Important Python libraries that are frequently used in the fields of scientific computing and data analysis include:

1. Pandas
2. NumPy
3. SciPy

In order to handle, manipulate, and analyze massive datasets, these libraries offer effective data structures and cutting-edge technologies, empowering researchers and scientists to gain insightful knowledge from their data.

Pandas is a flexible data analysis and manipulation package that provides solid data structures like DataFrames and Series. DataFrames are two-dimensional data structures that resemble tables and have columns that can store different kinds of data. In contrast, series are one-dimensional arrays that may store heterogeneous data. Data cleaning, filtering, transformation, aggregation, and merging processes are made easier by a comprehensive collection of functions and techniques offered by Pandas. Researchers may effectively extract and alter particular subsets of data by utilizing the potent indexing and slicing capabilities of Pandas, enabling exploratory data analysis and preprocessing tasks. A smooth data sharing and interoperability are also made possible by Pandas' seamless integration with other scientific libraries.

In scientific computing, NumPy, often known as Numerical Python, is crucial. The multidimensional array object known as the ndarray, which effectively stores and manipulates homogeneous data, is introduced. Compared to standard Python lists, NumPy arrays offer a number of benefits, including improved memory economy and mathematical operation optimization. When working with complex numerical computations, this capability is especially crucial.

SciPy offers a wide array of techniques for resolving optimization issues, enabling researchers to identify the best solutions for intricate systems. The numerical integration module simplifies operations involving single and double integration, the solution of differential equations, and numerical integration methods. Researchers may create new data points within the range of existing data using interpolation techniques included in SciPy, which makes accurate analysis and modeling possible. SciPy's support for linear algebra includes fundamental operations such matrix manipulation, decomposition, and linear system solution. These features are crucial for mathematical modeling and scientific simulations.

Tasks like noise reduction, feature extraction, and picture enhancement are made easier by the signal and image processing modules in SciPy, which provide a wide range of methods for analyzing, filtering, and manipulating signals and images. Researchers have access to a variety of statistical functions, probability distributions, and hypothesis testing tools thanks to SciPy's statistics package. In scientific research, these instruments are essential for statistical analysis, hypothesis validation, and decision-making. Additionally, SciPy includes extra modules for dealing with sparse matrices, geographic data structures, optimization with constraints, as well as several other scientific computing applications.

Scientists and researchers may manage, analyze, and understand scientific data more effectively by combining the strengths of Pandas, NumPy, and SciPy. These libraries equip researchers with the tools necessary to manage intricate datasets, carry out difficult mathematical operations, and conduct advanced statistical analysis, eventually enabling ground-breaking scientific achievements.

4.1.2 Loading the data

```
1 df = pd.read_excel("ADNIMERGE_R_220706.xlsx")
```

Listing 4.2: Importing the libraries in Python

The line of code reads the 'ADNIMERGE_R_220706' Excel file using the Pandas library and stores its information in a DataFrame called df. The Pandas library, which was imported previously in the code, is referred to as pd. An Excel file may be read using the Pandas function read_excel(). Read_excel() is called by the function read_excel('ADNIMERGE_R_220706.xlsx'). You provide the name of the Excel file you wish to read between parenthesis. The variable df is given the read_excel() response's return value. By custom, the DataFrame object in Pandas is commonly referred to by the abbreviation df. The method opens the 'ADNIMERGE_R_220706.xlsx' Excel file and reads its contents, storing them in a DataFrame so that you may manipulate, analyze, and perform various operations on the data using Pandas' robust features.

4.1.3 Filtering the DataFrame

Now that the ADNI DataFrame has been created, some filtering is needed.

Removing NaN's and converting string to numeric

```
df['ABETA.bl'] = df['ABETA.bl'].str.replace("<","")
df['ABETA.bl'] = df['ABETA.bl'].str.replace(">","")
df['ABETA.bl'] = pd.to_numeric(df['ABETA.bl'])
```

Listing 4.3: Removing NaN's and converting string to numeric

Filter only the amyloid pathology patients

```
filtered_df = df[(df["DX"] == dx_category) & (df["VISCODE"] == "bl") & (df["PTAU.bl
    "]/df["ABETA.bl"] > 0.028)]
```

Listing 4.4: Filter only the amyloid pathology patients

The code essentially creates a subset of the original DataFrame df that includes only the rows where the "DX" column matches "CN", "MCI" or "Dementia", the "VISCODE" column corresponds to the "baseline" time point, and the ratio of "PTAU.bl" to "ABETA.bl" is greater than 0.028. This filtering approach aims to identify individuals with amyloid pathology, as indicated by the chosen threshold value of 0.028 for the Pτ to Aβ ratio.

4.2 Processing the data

After the preparation of the data now comes the most important part which is to process the data.

4.2.1 Categorical and numerical data

As data can be represented in two ways namely categorical and numerical, I will also process the data in two ways with the first one being categorical and the second one numerical.

Categorical data

Here is an example of how categorical data such as gender can be extracted from the DataFrame for obtaining information about the different Alzheimer groups:

```
#FEMALE
female_subjects = filtered_df[filtered_df["PTGENDER"] == "Female"]["RID"].nunique()
female_percentage = f"{(female_subjects/unique_subjects)*100:.1f}%"
```

Listing 4.5: processing categorical data

filtered_df[filtered_df["PTGENDER"] == "Female"] filters the filtered_df DataFrame to include only rows whose value in the "PTGENDER" column is "Female". This condition selects rows corresponding to female subjects. ["RID"].nunique() executes the nunique() function on the "RID" column of the filtered data frame, which calculates the number of unique subject identifiers ("RID") among the selected rows. This yields the number of female subjects. The resulting number of female subjects is assigned to the variable female_subjects. (female_subjects/unique_subjects)*100 calculates the percentage of female subjects by dividing the count of female subjects (female_subjects) by the total number of unique subjects in the dataset (unique_subjects). This percentage is then multiplied by 100. f"(female_subjects/unique_subjects)*100:.1f%"

formats the calculated percentage value as a string with one decimal followed by a percent symbol ("%"). This ensures that the percentage is displayed with one decimal place.

Numerical data

For the numerical data, the built in methods from NumPy is used to obtain information such as mean, min, max, standard deviation and much more. In the code listing below, it can be seen how these methods work to obtain information about Age of a given group.

```
#AGE
age_mean = filtered_df["AGE"].mean()
age_std = filtered_df["AGE"].std()
age_min = filtered_df["AGE"].min()
age_max = filtered_df["AGE"].max()
age_stats = f"{age_mean:.1f} ({age_std:.1f}, {age_min:.1f}, {age_max:.1f})"
```

Listing 4.6: processing numerical data

4.2.2 Statistical Analysis

Performing statistical analysis is crucial to get information between the groups and about the groups themselves. One way ANOVA is performed with Turkey Post-Hoc correction to analyse the significant differences between groups. In the code below can be seen how a Python function is created with as input the three groups which would be liked to analysed.

```
def one_way_anova_with_posthoc(group1, group2, group3):
    f_value, p_value = f_oneway(group1, group2, group3)
    print(p_value)
    if p_value < 0.01:
        data = np.concatenate([group1, group2, group3])
        labels = ['group1'] * len(group1) + ['group2'] * len(group2) + ['group3'] *
 len(group3)
        tukey_results = pairwise_tukeyhsd(data, labels, 0.01)
        print(tukey_results)
```

Listing 4.7: Function for statistical analysis

The one-way ANOVA test is used by the code to first get the F-value and p-value. The p-value was then printed. A post-hoc Tukey's test is run by the code if the p-value is less than 0.01. The data from all groups are pooled into a single array for the Tukey's test, and labels are made to distinguish the groups. To ascertain if there are significant differences between groups, the pairwise Tukey's test is used. The outcomes are then reported, together with the pairwise comparisons and modified p-values. In short, the code enables statistical analysis to assess group differences using one-way ANOVA and post-hoc Tukey's test, giving information about the importance of the observed variances across various groups.

4.2.3 Longitudinal Data

To analyse the evolution of Alzheimer's Disease of the patients it is also crucial to perform longitudinal analysis. With the code below we get the maximum value for the Pτ over Aβ value for each row, sorting the this column from large to small and afterwards dropping all the duplicates at the 'RID' column gives us the maximum for each subject for this value and thus longitudinal data. The code does this for

the CN group and also shows how the T+ and T- of the AT(N) classification gets derived regarding the treshold value.

```
filtered_CN_Long = filtered_CN1.sort_values('PTAUABETA', ascending=False).
    drop_duplicates(['RID'])
CN_Long_TMinus = filtered_CN_Long[filtered_CN_Long["PTAU"]<=27]["RID"].nunique()
CN_Long_TPlus = filtered_CN_Long[filtered_CN_Long["PTAU"]>27]["RID"].nunique()
```

Listing 4.8: Longitudinal Data

4.3 Machine Learning Models

Four Machine Learning models are created, two models each for the preclinical and prodromal stage. Each stage has a model based on data from baseline and also based on longitudinal data.

4.3.1 Preparing the Cox Model

Event

First the cox model gets prepared to perform a machine learning model. In the code below it is shown for the prodromal stage but for the preclinical stage the exact same thing happens.

```
progressed = MCI[MCI['DX'].isin(['Dementia'])]
# Filter out the subjects who have progressed to Dementia
help_df = MCI[MCI['DX'].isin(['Dementia'])]['RID']
stable = MCI[~MCI['RID'].isin(help_df)]
# Give all progressed subjects a 1 for 'progressed' column and 0 for the stable
    subjects
stable["event"]=0
progressed["event"]=1
```

Listing 4.9: Preparing the Cox Model: Event column

Line 1 of the code filters out participants who have advanced to dementia from the MCI DataFrame. It then generates a new DataFrame named "progressed" that only contains people whose "DX" column has the value "Dementia.". New DataFrame 'help_df' is created on line 2. It records the 'RID' (subject ID) values of the subjects who have advanced to dementia precisely. The 'RID' column is taken out of the 'progressed' DataFrame. The new DataFrame'stable' is established on line 3. It contains all of the MCI DataFrame individuals who have not yet developed dementia. The participants whose 'RID' values are absent in 'help_df', suggesting that they have not advanced to dementia, are excluded using the 'help_df' DataFrame. The'stable' DataFrame has a new column named 'event' inserted on line 5. This column has a value of 0 for each row. This column serves to demonstrate that the participants are stable and have not developed dementia. The 'event' column is lastly added to the 'progressed' DataFrame at line 6. Similar to the preceding line, a value of 1 is assigned to each row in this column. Now, the 'event' column shows that these people have advanced to dementia.

Duration

```
stable.sort_values('Month.bl', inplace=True)
# Drop duplicates based on 'RID' column, keeping the last (latest) observation:
    Stable keeping only the highest value for censoring time
```

```
3 stable.drop_duplicates(subset='RID', keep='last', inplace=True)
4 stable['duration'] = stable['Month.bl']
5
6 #Progressed calculating conversion time, first time MCI becomes Dementia
7 baseline_month = progressed.loc[progressed['DX'] == 'Dementia', 'Month.bl'].min()
8 progressed['duration'] = progressed['Month.bl']
9 progressed.drop_duplicates(subset="RID", inplace=True)
10 #Merge the two DataFrames into one for Cox Model
11 MCI_merged = pd.concat([progressed, stable])
```

Listing 4.10: Preparing the Cox Model: Duration column

The values in the 'Month.bl' column are used to sort the'stable' DataFrame in ascending order on line 1. This sorting process is in-place, which means that it changes the'stable' DataFrame right away. Line 3 of the code removes duplicate entries from the "stable" DataFrame based on the "RID" column. For each distinct "RID" value, only the most recent observation is retained; all prior duplicate rows are discarded. This procedure immediately alters the "stable" DataFrame thanks to the 'inplace=True' argument.

The values for the new column "duration," which is added to the "stable" DataFrame in line 4, are taken from the "Month.bl" column. The month values from the "Month.bl" column in the "stable" DataFrame are captured in a new column called "duration" that is created in this phase. This ultimately leads to the censoring time of the stable patients. Line 7 of the code is where the baseline month for the "progressed" DataFrame is determined. It pulls the lowest value from the "Month.bl" column and chooses rows from the "progressed" DataFrame where the "DX" (diagnostic) column is "Dementia." The MCI (Mild Cognitive Impairment) individuals' first month of dementia progression is represented by this baseline month. This ultimately leads to the conversion time of the progressed patients, because the code filters only the rows where the MCI patients become Dementia patients for the first time, hence conversion time. The values for the new "duration" column, which is added to the "progressed" DataFrame in line 8, are taken from the "Month.bl" column. In the 'progressed' DataFrame, a new column called 'duration' is now created, just as it was in the 'stable' DataFrame. Line 9 of the code removes duplicate rows from the "progressed" DataFrame based on the "RID" column, only maintaining the first instance of each distinct "RID" value.

Finally, in line 11, the 'progressed' and 'stable' DataFrames are concatenated using the 'pd.concat' function. This operation merges the two DataFrames into one, creating a new DataFrame called 'MCI_merged'. The resulting DataFrame contains both the 'progressed' and 'stable' subjects, allowing for further analysis, such as applying a Cox Model.

Standardising the values of the combination vector

```
1 stable.sort_values('Month.bl', inplace=True)
2 from scipy.stats import zscore
3 MCI_merged.dropna(subset=["ADAS13.bl"], inplace=True)
4 covariates = ['RAVLT.learning.bl', 'RAVLT.perc.forgetting.bl', 'ADAS13.bl', 'CDRSB.
      bl', 'LDELTOTAL.bl']
5
6 # Create a new DataFrame with standardized values
7 MCI_standardized = MCI_merged.copy()
8
9 MCI_standardized[covariates] = MCI_standardized[covariates].transform(zscore)
10 # Only taking the relevant columns for the cox modeling
```

```
cox_column = ['event', 'duration', 'RAVLT.learning.bl', 'RAVLT.perc.forgetting.bl',
    'ADAS13.bl', 'CDRSB.bl', 'LDELTOTAL.bl']

MCI_finalised = MCI_standardized[cox_column]
# Getting rid of the 0 duration rows as this could harm the modeling of the cox
    model
MCI_finalised = MCI_finalised[MCI_finalised["duration"] != 0]
```

Listing 4.11: Preparing the Cox Model: Standardising the used covariates columns

The best combination of biomarkers is selected, according to studies by the supervisor, prof. Carlos Platero. For Preclinical:

$$RAVLTLearning, FAQ, ADAS13, CDRSB, MMSE, AGE$$

For Prodromal:

$$ADAS13, FAQ, MMSE, CDRSB.bl$$

In the 'MCI_merged' DataFrame, missing values are first addressed by removing rows where the 'ADAS13.bl' column contains NaN values. This makes that the dataset is still comprehensive and prepared for additional research. After that, a list named "covariates" is made with the names of the columns that will be standardized. These columns are chosen depending on how well they apply to the next modeling job. The 'MCI_merged' DataFrame is copied into a new DataFrame with the name 'MCI_standardized'. This process enables us to maintain the original data while working with standardized numbers. The z-score transformation is then used to normalize the chosen columns in the 'covariates' list.

With the help of this procedure, it is guaranteed that the values in these columns are scaled to have a mean of 0 and a standard deviation of 1. When multiple features have distinct scales, standardization is advantageous because it harmonizes the scales and prevents any one aspect from predominating the analysis. A list named "cox_column" is made with the essential columns, such as "event," "duration," and other pertinent factors, to prepare the data particularly for Cox modeling. By choosing the appropriate columns from the "MCI_standardized" DataFrame, a new DataFrame called "MCI_finalised" is produced using the "cox_column" list. By keeping only the columns that are important for the Cox modeling analysis, this phase streamlines the dataset. The 'MCI_finalised' DataFrame is then filtered to remove any entries where the 'duration' column equals 0. This guarantees that the final dataset only contains subjects with a non-zero duration, which is necessary for accurate modeling.

The algorithm prepares the data for later analysis, such as Cox modeling, by executing these data pretreatment and feature selection stages, making sure that missing values are handled, pertinent columns are standardized, and the dataset is suitably prepared for additional research.

Longitudinal data

For longitudinal data we take the DataFrame as extracted in the same way like previous section and we also perform the LME Model to prepare the COX model. This code uses a linear mixed-effects (LME) methodology to estimate the course of prodromal Alzheimer's disease. To get longitudinal data ready for Cox modeling.

Duration is the dependent variable in the LME model, while the independent variables ADAS13, FAQ, MMSE, and CDRSB are measurements or evaluations of the progression of Alzheimer's disease.

```
# Specify the LME formula
lme_formula = 'duration ~ ADAS13 + FAQ + MMSE + CDRSB'
# Fit the LME model
lme_model = smf.mixedlm(lme_formula, data= data, groups=data['RID'])
lme_result = lme_model.fit()
random_residuals=lme_result.resid
data['random_residuals']=random_residuals
```

Listing 4.12: LME Model

4.3.2 Cross-validation with k-fold

Cross-validation and the Cox proportional hazards (CoxPH) model for survival analysis are implemented in the given code section. Using this method, we can analyze the model's performance on hypothetical data and determine its propensity to forecast survival outcomes.

```
from sklearn.model_selection import KFold
from lifelines import CoxPHFitter
# Split into test set and model data set
model_data, test_data = train_test_split(MCI_finalised, test_size=0.1, random_state
    =42)
# Perform 10-fold cross-validation
kfold = KFold(n_splits=10, shuffle=False)
# Initialize an empty list to store performance metrics
performance_metrics = []
for train_index, eval_index in kfold.split(model_data):
    # Split data into train and evaluation sets for the current fold
    train_data = model_data.iloc[train_index]
    eval_data = model_data.iloc[eval_index]
    # Initialize a new Cox model
    cox_model = CoxPHFitter()
    # Fit the Cox model on the train data
    cox_model.fit(train_data, duration_col='duration', event_col='event')
    # Evaluate the performance of the model on the evaluation data
    c_index = cox_model.score(eval_data, scoring_method="concordance_index")
    # Store the performance metric for this fold
    performance_metrics.append(c_index)
# Calculate the average performance across all folds
average_performance = sum(performance_metrics) / len(performance_metrics)

# Print the average performance
print("Average C-index: ", average_performance)
# Fit the Cox model on the entire training dataset
cox_model.fit(model_data, duration_col='duration', event_col='event')
# Predict on the test dataset
predictions = cox_model.predict_median(test_data)
# Evaluate the model's performance on the test dataset
c_index_test = cox_model.score(test_data, scoring_method="concordance_index")
print(c_index_test)
```

Listing 4.13: Progression model with cross-validation

The necessary libraries are first imported, including CoxPHFitter from the lifelines package and KFold from the sklearn.model_selection module. The train_test_split function is then used to divide the dataset into a model dataset and a test dataset. The test dataset will be utilized for the final performance evaluation, whereas the model dataset will be used for training and evaluation. A 10-fold strategy is built up using KFold with 10 splits to do cross-validation. The procedure does not involve shuffling the data.

To keep the performance metrics collected during cross-validation for each fold, a blank list called "performance_metrics" is made. Using the indices received by kfold.split(), the model dataset is further split into training and evaluation sets for

the current fold throughout the cross-validation cycle. CoxPHFitter() is used to set up a fresh instance of the CoxPH model. Utilizing the fit technique and the duration and event columns, the Cox model is fitted (trained) on the training data. The concordance index (c_index) is used as the performance indicator to assess the performance of the Cox model using the evaluation data. Utilizing the Cox model's scoring technique, this is accomplished.

The 'performance_metrics' list is supplemented with the acquired performance metric (c_index) for the current fold. The average performance across all folds is determined when the cross-validation cycle has been completed by adding the performance measurements and dividing by the number of folds. The average concordance index attained by the Cox model during cross-validation is represented by the printed value of average performance (average_performance). In order to use all available data for training, the Cox model is then fitted on the whole model dataset using the fit technique. The predict_median technique of the Cox model is used to create predictions on the test dataset. The concordance index (c_index_test) is used as a performance statistic to assess the model's performance on the test dataset. The concordance index for the test dataset is then reported, giving information on how well the model performs on unobserved data.

4.3.3 Performance Evaluations and determining the duration of the preclinical and prodromal stages

```
# Calculate the average performance across all folds
average_performance = sum(performance_metrics) / len(performance_metrics)
# Calculate the standard deviation of the performance metrics
std_dev = np.std(performance_metrics)
# Calculate the confidence interval (95% confidence level)
confidence_interval = stats.t.interval(0.95, len(performance_metrics) - 1, loc=
    average_performance, scale=std_dev / np.sqrt(len(performance_metrics)))
# Print the performance
print("Average C-index: ", average_performance)
print("Confidence Interval: ", confidence_interval)
# Concatenate all fold predictions
all_predictions = pd.concat(fold_predictions)
# Calculate the average survival lifetime with confidence intervals
median_survival_lifetime = all_predictions.median()
lower_ci = all_predictions.quantile(0.05)
upper_ci = all_predictions.quantile(0.95)
# Print the average survival lifetime and confidence intervals
print("Median Survival Lifetime: ", median_survival_lifetime/12)
print("Survival Time Confidence Intervals: ({}, {})".format(lower_ci/12, upper_ci
    /12))
# Fit the Cox model on the entire training dataset
cox_model.fit(model_data, duration_col='duration', event_col='event')
# Predict on the test dataset
predictions = cox_model.predict_median(test_data)
# Evaluate the model's performance on the test dataset
c_index_test = cox_model.score(test_data, scoring_method="concordance_index")
print("Concordance Index when testing the model on test data:", c_index_test)
```

Listing 4.14: Performance Evaluation

The average performance across all folds is determined in the first section of the code. The average is determined by adding together all the performance metrics in a list and dividing the total by the number of metrics. This list is saved in the variable performance_metrics. The performance metrics' standard deviation must then be determined. This is accomplished by measuring the spread or variability of the performance metrics around the average using the np.std() function from the NumPy

library. The algorithm then determines a confidence interval with a 95% degree of assurance. It makes use of the SciPy library's stats.t.interval() function, which takes into consideration average performance, standard deviation, and a variety of performance indicators.

The average C-index and the confidence interval are then printed by the code. The pd.concat() method from the pandas library is then used to concatenate every fold prediction that is stored in the fold_predictions variable. The variable all_predictions is given the concatenated predictions. The median survival lifespan is then calculated using the all_predictions dataframe in the next step. The lower and higher confidence intervals for the survival time are then determined by the code. The 5th and 95th percentiles are obtained by using the quantile() function on all_predictions with the inputs 0.05 and 0.95, respectively. Lower_ci stands for the lower confidence interval, and upper_ci for the higher confidence interval. The code, after fitting the model, uses the trained cox_model and the predict_median() function to forecast the median survival time for the test dataset. The predictions variable is used to store the outcomes as predictions. The concordance index (c-index), which is determined using the score() function on cox_model, is the last step in the code's evaluation of the model's performance on the test dataset. The scoring_method option is set to "concordance_index" and the test_data are given as input. The c_index_test variable holds the outputted c-index, which is subsequently printed. For the test dataset, predictions are made using the Cox model's predict_median function. The median survival time for each person in the dataset is predicted by this approach. The variable predictions contains the forecasts. The median expected survival time is computed by locating the projections' median value in order to understand the findings. Assuming the survival time is expressed in months, the estimatingated median survival time is divided by 12 to translate it to years. The median survival time is then produced, giving an estimate of how long it will likely be until the event happens for the people in the test dataset. Similarly, by determining the mean value of the forecasts, the mean expected survival time is determined. To convert the mean survival time from months to years, divide it by 12. The average time until an event happens for the people in the test dataset is then estimated using the mean survival time, which is printed. These calculations shed light on the model's expectations for the duration of predicted survival for each person in the test dataset.

4.3.4 Kaplan-Meier

In survival analysis, the Kaplan-Meier estimator is a statistical technique used to calculate the survival function of a population that has experienced an interesting event over time. [37] To evaluate and show survival data, it is frequently used in medical research, notably in the field of cancer. The survival function, sometimes referred to as the survival probability or survivor function, calculates the likelihood that a person or a group will live through a specific point in time. In other words, it determines the percentage of those who have not gone through the event (such as death or the growth of a disease) by a certain period. When working with censored data, which is when the event of interest has not happened for any people during the study period, the Kaplan-Meier estimator is especially helpful. In this case for the preclinical and prodromal stage the event is the fact that a subject converts or

not to respectfully MCI and AD. Censoring can occur for a number of reasons, for as when participants are lost to follow-up or when the research is over before the event occurs for everyone involved. The estimator gives a non-parametric prediction of the survival function while taking into consideration these censored observations. The Kaplan-Meier estimator uses all available data points until an event occurs or censoring occurs, which is one of its main benefits. It can also accommodate partial or unequal follow-up intervals. It offers estimates of survival probability at various time points and enables the examination of time-to-event data in the context of censoring. The estimator also makes it possible to compare survival curves between various groups or subpopulations. This enables scientists to determine if various elements, such available treatments or patient characteristics, have an effect on survival rates. The Kaplan-Meier estimator also offers Kaplan-Meier plots as a means of visualizing the survival curves. These charts show the predicted survival probabilities with time, making it simple to understand and evaluate survival trends. In conclusion, the Kaplan-Meier estimator is an effective tool for calculating the survival function and examining time-to-event data in survival analysis. It is useful in clinical investigations and medical research since it accepts censored observations and makes it easier to compare survival curves between various groups. [36]. Below you can see the implemented code for performing Kaplan-Meier analysis.

```
from lifelines import CoxPHFitter, KaplanMeierFitter
# Fit the Kaplan-Meier model on the test data
kmf = KaplanMeierFitter()
kmf.fit(test_data['duration'], test_data['event'])

# Plot the Kaplan-Meier survival curve for the test dataset
kmf.plot()
plt.title('Kaplan-Meier Survival Curve (Test Data)')
plt.xlabel('Time (months)')
plt.ylabel('Survival Probability')
plt.show()
```

Listing 4.15: Performing Kaplan-Meier

Chapter 5

Results

The results of the research undertaken in this thesis are presented in this part. The goal of this study was to use Artificial Intelligence techniques to assess the temporal length of the preclinical and prodromal phases of Alzheimer's disease in individuals with amyloid pathology. The acquired data was examined using a combination of statistical and machine learning techniques. The study's findings will be detailed in the subsections that follow, including an analysis of the evolution of Alzheimer's disease in patients, the tau pathology and neuro-degeneration profiles of the study subjects, the biomarker predictors with the greatest power to estimate the evolution of Alzheimer's disease, and an estimation of the duration of the preclinical and prodromal stages of the disease.

Each subsection will begin with a description of the findings, followed by an in-depth assessment of their importance, relevance to the current literature, and resolution of any limits or discrepancies. Furthermore, the ramifications of the findings will be examined in terms of their theoretical and practical contributions, and prospective future study directions will be suggested. Overall, the purpose of this part is to give a thorough analysis of the data obtained as well as insights into the research issues addressed in this thesis.

5.1 Table 1: Baseline demographic and clinical characteristics of the subjects.

The subjects' baseline demographic and clinical characteristics. Data are expressed as mean (SD) (minimum, maximum) or n(%). For baseline age, education, neuropsychological score, and AD CSF biomarkers, ANOVA with Turkey post hoc test is used, except for gender and APOE4 convert, where the chi-square test is used. A p-value of less or equal than 0.01 is considered statistically significant. In the Post-Hoc column the a, b and c mean respectively significance detected between CU and MCI, CU and AD and between MCI and AD. See table 5.1.

Gender and APOE4 The percentage of females in the CN group was significantly greater (56.3%) than in the MCI (39.1%) and AD (43.6%) groups (p¡0.01). Furthermore, the APOE4 genotype was substantially more common in the MCI (70.9%) and AD (74.0%) groups than in the CN (55.2%) group. There were no significant differences between the MCI and AD groups.

Age and Education The individuals' mean age did not differ substantially between the MCI and AD groups, however both were considerably younger than

Group	CN	MCI	Dementia	Post-Hoc
Subjects	87	320	204	
Female	49 (56.3%)	125 (39.1%)	89 (43.6%)	a
APOE4	48 (55.2%)	227 (70.9%)	151 (74.0%)	a,b
Age	76.2 (5.2, 65.4, 87.2)	73.5 (7.1, 54.4, 87.8)	74.1 (8.1, 55.6, 90.3)	a
Education	16.0 (2.7, 6.0, 20.0)	15.9 (2.8, 6.0, 20.0)	15.4 (2.9, 4.0, 20.0)	b
MMSE	29.1 (1.1, 24.0, 30.0)	27.2 (1.9, 23.0, 30.0)	23.2 (2.0, 19.0, 27.0)	a,b,c
CDRSB	0.05 (0.16, 0.00, 1.00)	1.67 (0.93, 0.50, 5.00)	4.44 (1.64, 1.00, 10.00)	a,b,c
ADAS13	10.3 (4.4, 2.0, 24.0)	19.1 (6.7, 4.0, 38.0)	30.3 (8.0, 14.0, 51.0)	a,b,c
FAQ	0.4 (0.8, 0.0, 3.0)	4.0 (4.4, 0.0, 22.0)	12.9 (6.7, 0.0, 28.0)	a,b,c
Aβ	687.7 (256.9, 200.0, 1700.0)	652.1 (201.8, 210.9, 1700.0)	592.4 (199.4, 212.3, 1700.0)	b,c
TAU	307.2 (105.4, 109.7, 590.1)	361.6 (140.0, 130.7, 1300.0)	380.3 (144.1, 145.9, 851.6)	a,b
pTAU	30.7 (11.3, 10.9, 60.0)	36.8 (15.3, 11.8, 120.0)	38.3 (15.5, 13.5, 94.9)	a,b
pTAU/Aβ	0.048 (0.023, 0.029, 0.180)	0.060 (0.028, 0.028, 0.237)	0.069 (0.031, 0.028, 0.245)	a,b,c
FDG	n = 71, 1.30 (0.12, 0.99, 1.67)	n = 258, 1.21 (0.13, 0.73, 1.70)	n = 161, 1.06 (0.13, 0.72, 1.40)	a,b,c

Table 5.1: The subjects' baseline demographic and clinical characteristics

the CN group. The mean education level of the CN and MCI groups did not differ substantially, however both groups had considerably higher education levels than the AD group.

Neuropsychological scores The table shows the mean scores for a variety of neuropsychological exams. The Mini-Mental State Examination (MMSE) scores in the MCI and AD groups were substantially lower than in the CN group. The Clinical Dementia Rating Scale-Sum of Boxes (CDRSB) scores in the MCI and AD groups were substantially higher than in the CN group. The Alzheimer's Disease Assessment Scale-Cognitive Subscale 13 items (ADAS13) scores in the MCI and AD groups were substantially higher than in the CN group. The MCI and AD groups scored considerably higher on the Functional Activities Questionnaire (FAQ) than the CN group.

AD CSF Biomarkers The mean levels of cerebrospinal fluid (CSF) biomarkers for Alzheimer's disease (ABeta, TAU, phosphorylated TAU (pTAU), and the pTAU/A ratio are also showed. The A levels in the AD group were considerably lower than in the CN and MCI groups. TAU and pTAU levels were substantially greater in the AD group than in the CN and MCI groups. The pTAU/A ratio in the AD group was substantially greater than in the CN and MCI groups.

FDG The average results of the three groups' fluorodeoxyglucose (FDG) positron emission tomography (PET) scans are listed in table 1. In comparison to the CN group and the MCI group, the FDG levels in the AD group were respectively considerably lower and higher.

Conclusion Important demographic and clinical data from the participants in the CN, MCI, and AD groups are summarized in the table. The findings reveal substantial differences between the groups in terms of gender, APOE4, age, education, neuropsychological outcomes, AD CSF biomarkers, and FDG PET scan results. These findings may have an impact on clinical diagnosis and treatment and are important for comprehending the variability of cognitive impairment.

Feature	sCU	pCU	sMCI	pMCI
Subjects	55	32	135	185
Female	37 (67.3%)	12 (37.5%)	49 (67.3%)	76 (41.1%)
APOE4	26 (47.3%)	22 (68.8%)	94 (69.6%)	133 (71.9%)
Age	75.9 (5.6, 65.4, 87.2)	77.2 (4.3, 69.9, 85.6)	73.7 (6.6, 54.4, 87.8)	72.6 (7.2, 55.0, 87.7)
Education	15.7 (3.1, 6.0, 20.0)	15.6 (2.7, 12.0, 20.0)	15.8 (2.7, 10.0, 20.0)	15.9 (2.8, 6.0, 20.0)
MMSE	29.2 (1.0, 26.0, 30.0)	28.9 (1.4, 24.0, 30.0)	27.7 (1.8, 24.0, 30.0)	26.8 (1.7, 23.0, 30.0)
CDRSB	0.05 (0.17, 0.00, 1.00)	0.01 (0.07, 0.00, 0.50)	1.29 (0.74, 0.50, 4.00)	1.94 (1.00, 0.50, 5.00)
ADAS13	8.9 (4.2, 2.0, 19.0)	12.0 (3.8, 6.0, 24.0)	15.6 (6.0, 4.0, 36.3)	21.3 (6.0, 7.0, 38.0)
FAQ	0.3 (0.7, 0.0, 3.0)	0.3 (0.7, 0.0, 3.0)	2.1 (3.3, 0.0, 16.0)	5.6 (4.7, 0.0, 22.0)
Aβ	712.0 (285.9, 203.0, 1700.0)	644.0 (217.5, 200.0, 1261.0)	680.9 (228.5, 210.9, 1700.0)	630.4 (189.0, 267.2, 1187.0)
TAU	306.4 (110.3, 109.7, 590.1)	299.5 (89.8, 134.5, 492.1)	306.4 (110.3, 109.7, 590.1)	368.8 (125.6, 130.7, 816.9)
pTAU	30.5 (11.5, 10.9, 60.0)	29.9 (9.9, 12.8, 52.3)	34.1 (14.2, 13.3, 120.0)	37.5 (14.0, 11.8, 92.1)
pTAU/Aβ	0.046 (0.018, 0.029, 0.120)	0.051 (0.027, 0.029, 0.180)	0.053 (0.025, 0.028, 0.197)	0.063 (0.027, 0.028, 0.237)
FDG	n = 45,1.3 (0.1, 1.0, 1.5)	n = 26,1.3 (0.1, 1.1, 1.7)	n=116,1.32 (0.08, 0.99, 1.46)	n=142,1.17 (0.12, 0.91, 1.47)

Table 5.2: Baseline demographic and clinical characteristics of the subjects.

5.2 Table 2: Baseline demographic and clinical characteristics of the subjects

The baseline demographic and clinical features of the four groups—sCU, pCU, sMCI, and pMCI—are shown in the table. Following is a breakdown of the participants by group: sCU (n=55), pCU (n=32), sMCI (n=135), and pMCI (n=185). The distribution of gender differs between the groups. In comparison to pCU (37.5%) and pMCI (41.1%), the proportion of female individuals in sCU (67.3%) and sMCI (67.3%) is higher. The distribution of APOE4 genotypes varies between the groups as well. A larger proportion of patients in the pCU (68.8%), sMCI (69.6%), and pMCI (71.9%) groups than in the sCU group (47.3%) contain the APOE4 allele. The groups with the highest mean ages are sCU (74.9 years) and pCU (77.2 years), whereas the groups with the lowest ages are sMCI (73.7 years) and pMCI (72.6 years).

The mean ages of the educational levels groups range from 15.6 to 15.9 years and are similar. Neuropsychological test mean scores indicate a pattern of declining scores from the sCU group to the pMCI group. For instance, sCU has the greatest MMSE score (29.2) and the lowest MMSE score (26.8) is from pMCI, although pMCI has the highest CDRSB and ADAS13 scores (1.94 and 21.3, respectively) and sCU has the lowest (0.05 and 8.9). The CSF biomarkers (A, TAU, pTAU, and pTAU/A) do not exhibit a consistent pattern amongst the groups. However, the pMCI group had the greatest mean values for TAU (368.8), pTAU (37.5), and pTAU/A (0.063) and the highest mean value for A (630.4).

FDG data are available for a subset of subjects (n=45, 26, 116, and 142 for sCU, pCU, sMCI, and pMCI, respectively). The mean FDG values show no clear trend across the groups.

5.3 Table 3: Percentages of the AT(N) profiles in the CU, MCI, and AD subjects at baseline using AD CSF biomarkers and FDG-PET.

The table shows the percentages of AT(N) profiles utilizing AD CSF biomarkers and FDG-PET in the CU, MCI, and AD individuals at baseline. The number of participants in each group and the proportion of people in each group who had a positive FDG-PET scan are also included in the table. See table 6.1.

CU At baseline and at long-term follow-up, there were 87 individuals in the CU group. 42.5% of subjects had an A+T- baseline, whereas 57.5% had an A+T+ baseline. 35.6% of individuals in the long-term follow-up scored A+T-, whereas 64.4% scored A+T+. 71% of those who had FDG-PET scans were A+T+. The baseline group included 45.9% of the A+T+ people with neurodegeneration negative status, meanwhile the long-term follow-up group included 28.3% of them. On the other hand, 12.8% of the A+T+ people with neurodegeneration-positive status and 13.3% of those with baseline status were in the group receiving long-term follow-up.

MCI At baseline and at long-term follow-up, 320 people were included in the MCI group. At the start of the study, 26.2% of subjects had an A+T-, whereas 73.8% had an A+T+. 23.4% of patients in the long-term follow-up were A+T-, whereas 76.6% were A+T+. 80.3% of those who had FDG-PET scans were A+T+. A+T+ people with a neurodegeneration-negative status made up 36.1% of the baseline participants and 28.5% of the long-term follow-up subjects. In comparison, 35.4% of the A+T+ people who had neurodegeneration-positive status were in the long-term follow-up group whereas 37.2% were in the baseline group.

AD At baseline and at long-term follow-up, 204 individuals were in the AD group. At the outset, 23.5% of participants had an A+T-, whereas 76.5 percent had an A+T+. 22.1% of patients in the long-term follow-up were A+T-, whereas 77.9% were A+T+. 74.5 percent of patients who had FDG-PET scans were A+T+. 8.0% and 7.2%, respectively, of the A+T+ patients with neurodegeneration negative status were included in the baseline and long-term follow-up groups. In comparison, 69.3% of the A+T+ people who had neurodegeneration-positive status were in the long-term follow-up group whereas 67.8% were in the baseline group.

Conlusion Important details about the AT(N) profiles of Alzheimer's disease at various phases, including CU, MCI, and AD, are shown in the table. The information on the amounts of beta-amyloid, tau, and neurodegenerative biomarkers in each group may be used to determine if Alzheimer's disease is present and how far along it is. Researchers and clinicians may find the data in the table helpful in creating and assessing Alzheimer's disease therapies.

5.4 Table 4: Comparison of some risk factors between sCU and pCU subjects in the study population.

The table compares some risk variables between study participants with sCU (slow cognitive decline) and pCU (preclinical cognitive decline). The research uses binary category and quantitative data to shed light on probable dividing characteristics between these two groups.

Subjects and age The sCU group had a little older mean age (74.9 years) than

Group	A+T-	A+T+	n (+FDG)	A+T- (+FDG)	A+T+ (+FDG)	A+T+(N-)	A+T+(N+)
CU BSL, n=87	42.5%	57.5%	71	40.8%	59.2%	45.9%	13.3%
CU LONG, n=87	35.6%	64.4%	71	35.2%	64.8%	28.3%	12.8%
MCI BSL, n=320	26.2%	73.8%	258	26.7%	73.3%	36.1%	37.2%
MCI LONG, n=320	23.4%	76.6%	258	23.3%	76.7%	28.5%	35.4%
Dementia BSL, n=204	23.5%	76.5%	161	24.2%	75.8%	8.0%	67.8%
Dementia LONG, n=204	22.1%	77.9%	161	23.0%	77.0%	7.2%	69.3%

Table 5.3: Percentages of the AT(N) profiles in the CU, MCI, and AD subjects at baseline using AD CSF biomarkers and FDG-PET

Factor	pCU	pCU (fast | slow)	sCU	sCU (fast | slow)
n	32	20 | 10	55	35 | 20
Age	77.2 (4.3)	78.4 (4.4) | 75.1 (3.6)	75.9 (5.6)	76.5 (5.8) | 74.5 (5.2)
Female	37.5%	40.0% | 40.0%	67.3%	62.9% | 75.0%
APOE4	68.8%	65.0% | 70.0%	47.3%	51.4% | 40.0%

Table 5.4: Comparison of some risk factors between sCU and pCU subjects in the study population.

the pCU group (77.2 years). The mean age of the fast pCU group (78.4 years) was greater than the slow pCU group (75.1 years), though, when the pCU group was further separated into fast and slow groups. The age distribution was quite narrow, as evidenced by the fact that the standard deviation for age was often low. This evidence implies that age may affect how quickly cognitive loss occurs, with older people perhaps experiencing a quicker decrease.

Female and APOE4 In terms of binary categorical factors, the sCU group had a greater percentage of females (67.3%) than the pCU group (37.5%). The percentage of females in the pCU group stayed consistent at 40.0% for both the fast and slow groups, which is an interesting finding. This implies that gender might not be a defining characteristic of fast and slow pCU. Additionally, the sCU group had a lower number of APOE4 carriers (47.3%) compared to the pCU group (68.8%). The percentage of APOE4 carriers remained lower in the fast pCU group (65.0%) than in the slow pCU group (70.0%) when the pCU group was split into fast and slow groups. This raises the possibility that APOE4 carrier status is linked to a quicker onset of cognitive deterioration in the preclinical phases.

Conlusion This table offers important details on a few risk variables linked to cognitive decline in its early and preclinical phases. The pace of cognitive decline may depend on factors such as age, gender, and APOE4 status, with older people and those who have the gene perhaps experiencing a quicker decline. To ascertain the direction and causation of these connections, more investigation is necessary.

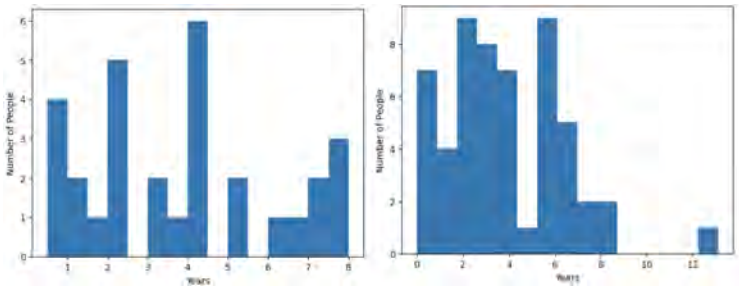

Figure 5.1: Convert and Censuring time (all subjects)

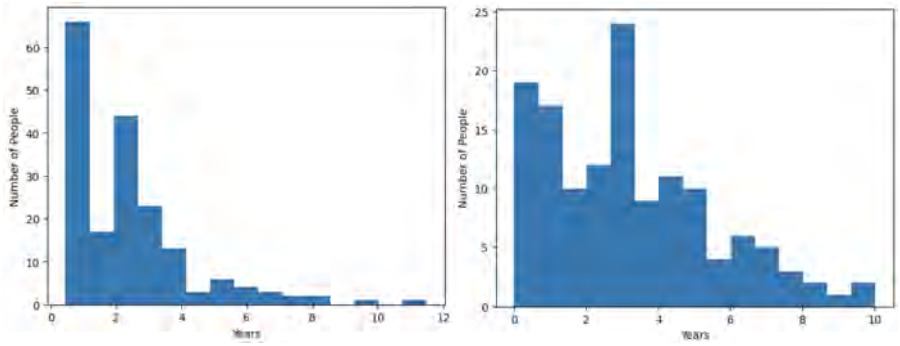

Figure 5.2: Convert and censuring time (all subjects)

5.5 Conversion rates of CU subjects towards MCI/dementia

Based on their conversion and censure time, Figure 6.1 shows the distribution of individuals into the pCU and sCU groups. Approximately 66.7% (n = 20/30) of all pCU-A+ patients, or about 22.9% (n = 20/87) of the CU-A+ population, converted within the first four years. Individuals who converted in the first four years make up around 12% of the total CU population of ADNI and about 45% of all pCU participants.

5.6 Conversion rates of MCI subjects towards dementia

The allocation of individuals into the two groups, pMCI and sMCI, according to their conversion or censuring time is shown in Figure 6.1. Within the first four years, around 50.6% (n = 162/320) of the MCI-A+ group, or nearly 89.0% (n = 162/182) of all pMCI cases, transitioned to dementia. Around 30% of MCI patients in the general ADNI sample transitioned to MCI during the first four years, while 80% of pMCI subjects converted to dementia within the same time period.

5.7 Percentages of the AT profiles in the CU, MCI and dementia subjects

A pTau/A ratio larger than 0.028 was utilized to fulfill the A+ requirement, while a pTau value greater than 27 pg/ml was used to establish T+. The baseline values are

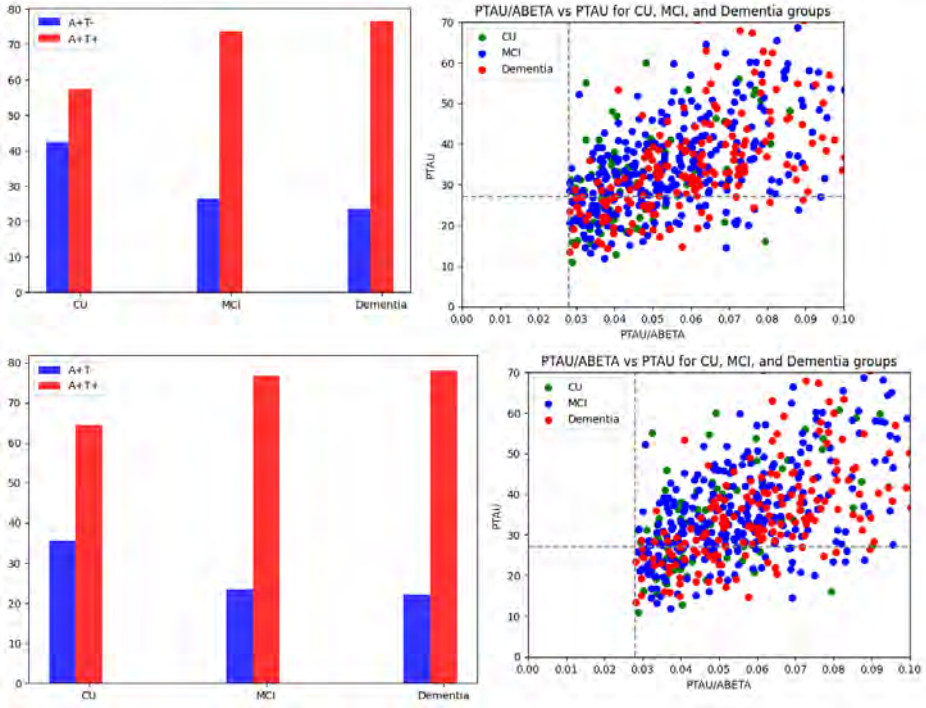

Figure 5.3: Percentages of the AT profiles in the CU, MCI and dementia subjects. Above is at baseline while the graphs below represent longitudinal analysis.

shown in the first column, and the longitudinal analysis' findings are shown in the second column. To show the measure cut-off positions, dashed lines were created. See Figure 5.7. The same conclusions are found as in table 6.1.

5.8 Correlation among AT(N) biomarkers at baseline

Figure 5.8 demonstrates a high linear correlation (R value more than 0.97) between Tau and pTau across all three clinical groups studied. This indicates that a T+ (pTau ¿ 27pg/ml) diagnostic would also suggest a N+ (Tau ¿ 290pg/ml) diagnosis. FDG PET and Tau in CSF are both diagnostic of neurodegeneration, with Tau in CSF showing the degree of neuronal damage at a certain moment and FDG representing cumulative neuropil loss and neuronal functional impairment. Although there is no significant association between Tau and FDG (with R values ranging from -0.06 to 0.08), there is considerable agreement when it comes to classifying neurodegenerative processes as positive (N+) or negative (N-). Approximately 70% of participants who tested positive for Tau also tested positive for FDG in the three clinical categories studied, but agreement for negative findings (N-) was poorer, falling below 40% for MCI and dementia cases and approximately 50% for CU subjects.

5.9 Machine Learning Models

After performing the code from the methods section, following results are obtained.

Preclinical Stage

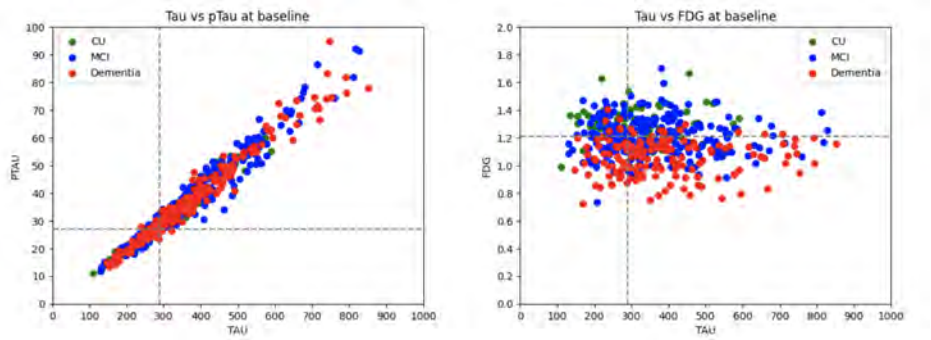

Figure 5.4: Percentages of the AT profiles in the CU, MCI and dementia subjects.

1. **Baseline**

   The average C-index calculated was 0.7133 (95% CI: 0.6937-0.7329), demonstrating a modest ability of the model to discriminate between the course of illness. This shows that the model may predict the chance of progression from the preclinical stage to a subsequent stage with some degree of accuracy. Individuals in the preclinical stage at baseline had an estimated median surviving lifespan of 7.98 years. This suggests that, on average, half of the people advanced to a new stage after around 8 years. The range of the median survival lifetime's 95% confidence interval, from 2.49 to 9.28 years, represents the likely range of the actual median surviving lifespan. When tested on test data, the model's concordance index was 0.64. This demonstrates that the model may classify individuals based on their likelihood of contracting a disease, albeit room for improvement still exists. In figure 5.9 you can see the Kaplan-Meier graph for preclinical stage.

2. **Longitudinal** The model's average C-index was 0.96, showing an excellent capacity for illness progression prediction which is close to perfection. With a 95% confidence range spanning 1.01 to 8.02 years, the projected median surviving lifespan for individuals was 4.07 years. The model's high concordance score of 0.89, which indicates good prediction accuracy in classifying people according to their likelihood of illness development, was evaluated using unknown data.

   However while these results (especially longitudinal) seem very promising and very high scoring, it is curcially important to note that this is only on one small dataset and further investigation is crucial. A possibility is also that some overfitting has happened.

Prodromal Stage

1. **Baseline**

   The results for the prodromal stage are a little better. The average C-index, which was 0.7582, showed that the model had a respectable capacity for discriminating. The range of the average C-index's 95% confidence interval, from 0.6887 to 0.8277, indicates where the real average C-index value is most likely to fall. Patients with MCI were found to have an estimated median surviving lifespan of 3.30 years. This is around the point at which half of the MCI patients are anticipated to advance to a new stage. The range of the 95 percent confidence interval for the median survival lifetime, which represents the likely

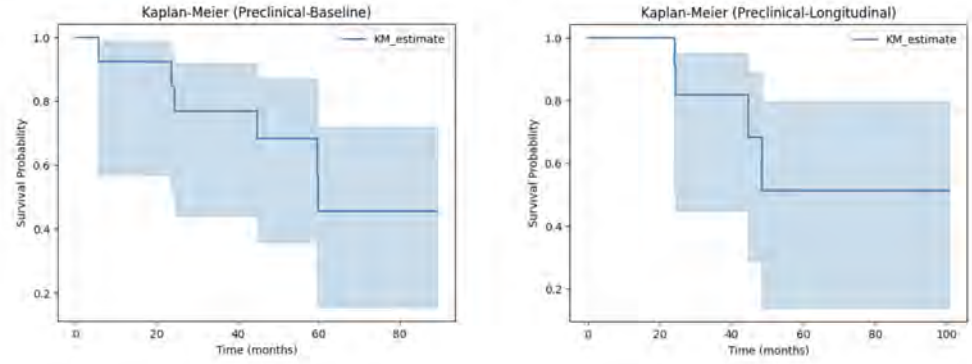

Figure 5.5: Kaplan-Meier graph for Preclinical stage

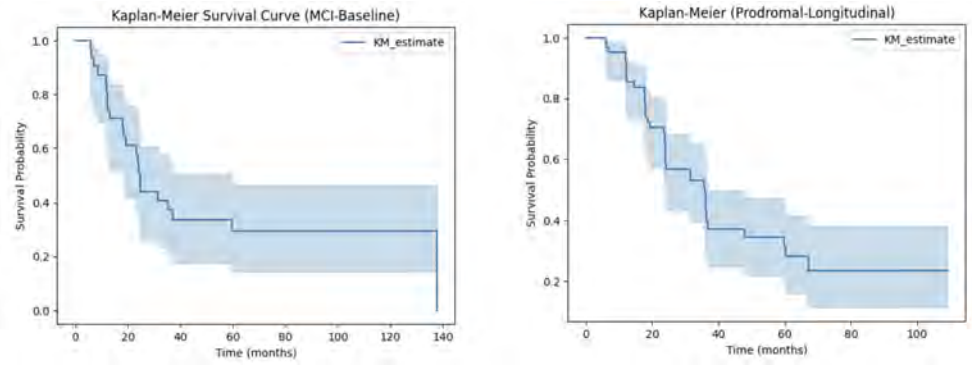

Figure 5.6: Kaplan-Meier graph for Prodromal stage

range of the actual median survival lifetime, was 0.88 to 9.99 years. The concordance index obtained while testing the model on the test data was 0.7322. This rating, which assesses the model's total predictive performance, shows a very feasible degree of prediction accuracy. In figure 5.9 you can see the Kaplan-Meier graph.

2. **Longitudinal**

   The longitudinal data model used for the prodromal period showed a strong capacity for prediction. The model's average C-index was 0.74053 (95% CI: 0.7168-0.7610), demonstrating a significant capacity for discrimination in determining the rate of illness development. This shows that the model is capable of efficiently differentiating between people who are most likely to advance to a later stage and those who could still be in the prodromal stage. In the prodromal stage, the projected median survivor lifespan was 3.30 years, indicating that, on average, half of the people advanced to a different stage during this time. The range of the median survival lifetime's 95% confidence interval, from 1.04 to 7.06 years, represents the likely range of the actual median survival lifespan. The model's concordance index was 0.70 when it was evaluated on hypothetical data. This index measures the model's overall prediction accuracy and shows how well it can rank people according to their probability of developing a disease.

Chapter 6

Discussion

6.1 Processing the data and feature engineering

The results section offers important details on the demographic and clinical baseline traits of the research subjects. For the development of progression ML models intended to determine the length of the prodromal and preclinical stages of Alzheimer's disease, the data supplied in this section are essential.

The baseline demographic and clinical features of the individuals are included in the analysis's first section. The distributions of APOE4 genotype and gender differ considerably amongst the groups. In comparison to the MCI and AD groups, the CN group had a larger proportion of females. Additionally, compared to the CN group, the MCI and AD groups have a higher prevalence of the APOE4 genotype. There are clear trends between the categories in terms of age and educational attainment. Although the mean ages of the MCI and AD groups are comparable, neither group is older than the CN group. The CN and MCI groups also have greater levels of education than the AD group. Scores on neuropsychological tests show significant disparities between the groups. In comparison to the CN group, the MCI and AD groups have lower MMSE scores and higher scores on the Clinical Dementia Rating Scale-Sum of Boxes (CDRSB) and the Alzheimer's Disease Assessment Scale-Cognitive Subscale 13 items (ADAS13). On the Functional Activities Questionnaire (FAQ), the MCI and AD groups perform better than the CN group. Aβ, TAU, phosphorylated TAU (pTAU), and the pTAU/Aβ ratio are only a few of the cerebrospinal fluid (CSF) indicators for Alzheimer's disease that are included in the study. In comparison to the CN and MCI groups, the AD group exhibits lower A levels and greater TAU and pTAU levels. Additionally, the AD group has a greater pTAU/Aβ ratio than the CN and MCI groups do. Additionally, the typical outcomes of positron emission tomography (PET) scans using fluorodeoxyglucose (FDG) are reviewed. In comparison to the CN group, the FDG levels in the AD group are noticeably lower, whereas those in the MCI group are greater. Understanding the neurodegenerative mechanisms linked to Alzheimer's disease is made easier by these results.

The percentages of AT(N) profiles in the CU, MCI, and AD participants at baseline utilizing AD CSF biomarkers and FDG-PET are also shown in the findings section. See table 6.1. The presence and development of Alzheimer's disease are assessed using these characteristics. The findings demonstrate changes in the AT(N) profiles between the groups, indicating various illness phases. Comparisons between

Group	A+T-	A+T+	n (+FDG)	A+T- (+FDG)	A+T+ (+FDG)	A+T+(N-)	A+T+(N+)
CU BSL, n=87	42.5%	57.5%	71	40.8%	59.2%	45.9%	13.3%
CU LONG, n=87	35.6%	64.4%	71	35.2%	64.8%	28.3%	12.8%
MCI BSL, n=320	26.2%	73.8%	258	26.7%	73.3%	36.1%	37.2%
MCI LONG, n=320	23.4%	76.6%	258	23.3%	76.7%	28.5%	35.4%
Dementia BSL, n=204	23.5%	76.5%	161	24.2%	75.8%	8.0%	67.8%
Dementia LONG, n=204	22.1%	77.9%	161	23.0%	77.0%	7.2%	69.3%

Table 6.1: Percentages of the AT(N) profiles in the CU, MCI, and AD subjects at baseline using AD CSF biomarkers and FDG-PET

subjects in prodromal cognitive aging and those with preclinical cognitive aging provide crucial insights into risk variables. The sCU group is marginally older than the pCU group and has a greater proportion of females, indicating differences in age and gender distribution across the groups. The proportion of APOE4 carriers varies depending on the group, with a smaller percentage in the sCU group than the pCU group. These results imply that the pace of cognitive deterioration in the preclinical stages of Alzheimer's disease may be influenced by age, gender, and APOE4 status. The conversion rates of CU participants to MCI/dementia and MCI subjects to dementia are next discussed. The information shows the proportion of people who converted within particular timeframes. These conversion rates offer important new information on the development of Alzheimer's disease from the preclinical to the clinical phases. Also the baseline correlations between the AT(N) biomarkers are investigated. According to the research, there is a strong linear correlation between Tau and pTAU, suggesting that a diagnosis of N+ is also suggested by a T+ diagnostic. Although Tau and FDG do not significantly correlate, there is wide agreement on whether neurodegenerative processes are positive (N+) or negative (-).

A very important result from the data preparation and data engineering part was the convert and censuring times of all subjects. This information is crucial to use to further analyse and comprehend the machine learning models such as which markers to use, what duration will be likely for both phases. For example if from these data we get a max duration of 10 years and the machine learning model predicts anything above this number we know there exists some trouble with the model and adjustment is needed and further investigation. The distribution of individuals into different groups based on conversion or censoring time is depicted in Figure 6.1. It reveals that within the first four years, approximately 66.7% (n = 20/30) of pCU-A+ patients, corresponding to about 22.9% (n = 20/87) of the entire CU-A+ population, underwent conversion. In the broader CU population of the ADNI study, these early conversions comprised approximately 12% of the total. Additionally, around 45% of all pCU participants experienced conversion within the first four years.

Figure 6.1 demonstrates the allocation of individuals into the pMCI and sMCI groups based on conversion or censoring time. Within the first four years, about 50.6% (n = 162/320) of the MCI-A+ group, or nearly 89.0% (n = 162/182) of all pMCI cases, transitioned to dementia. In comparison, around 30% of MCI patients

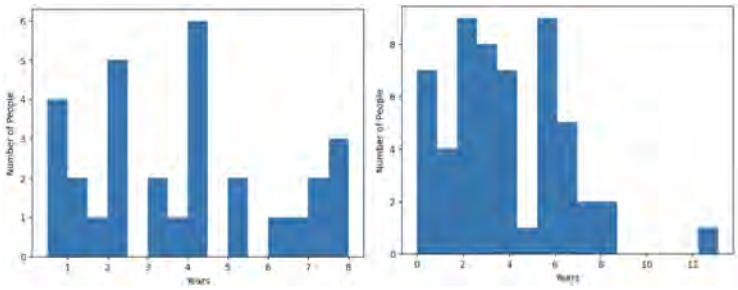

Figure 6.1: Convert and Censuring time (all subjects)

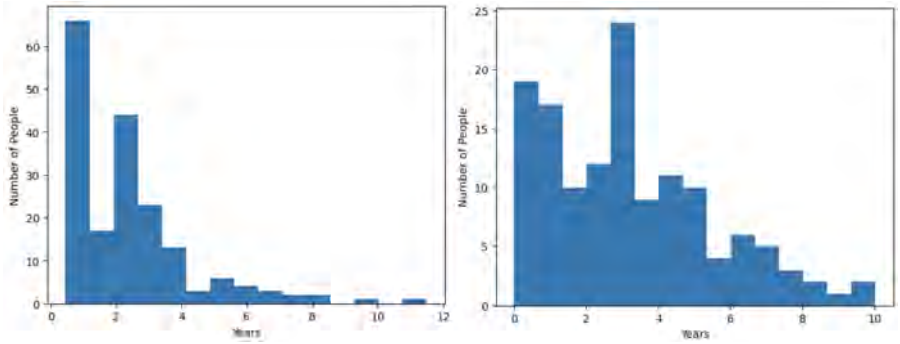

Figure 6.2: Convert and censuring time (all subjects)

in the general ADNI sample converted to MCI within the same timeframe, while 80% of pMCI subjects progressed to dementia. These findings emphasize the higher conversion rates and faster disease progression observed in the presymptomatic groups (pCU and pMCI) compared to the symptomatic groups (sCU and sMCI). The data provide insights into the timing and likelihood of conversion, contributing to our understanding of disease progression and its implications for management and intervention.

6.2 Machine Learning Models

In the discussion section the focus lies on understanding the results from the previous chapter.

Two models were divided into two categories: baseline and longitudinal, for each stage. The results provide insights into the predictive capacity of the models and their ability to discriminate between different stages of the disease. Both the baseline and longitudinal models showed some degree of accuracy in predicting the development to a future stage in the preclinical stage. The average C-index of 0.71 revealed that the baseline model had a good capacity to distinguish between disease progression. This shows that the model could be able to estimate the likelihood of advancement rather accurately. At baseline, the estimated median remaining lifetime of those in the preclinical stage was 7.98 years, meaning that, on average, half of the population moved on to a new stage after around 8 years. The model's concordance index on test data, which was 0.64, indicates that there is space for improvement in the model's categorization capabilities. With an average C-index of 0.96, the longitudinal model for the preclinical period demonstrated an excellent

ability to predict the course of the illness. Individuals were expected a survival time of 4.07 years, with a 90% confidence interval of 1.01 to 8.01 years. The model's strong concordance score of 0.89 on unidentified data shows great prediction accuracy in grouping individuals according to their propensity to get the disease. However it is once again very crucial to note that this is only for this particular dataset and maybe not applicable to the whole population, further investigation is crucial because overfitting could have been possible and the reason why the C-index for longitudinal is this high.

Data	**Average C-index**	**Survival Lifetime (90% CI)**	**C-index on test data**
Baseline	0.71	7.98 (4.03, 9.28)	0.64
Longitudinal	0.96	4.07 (1.01, 8.02)	0.89

Table 6.2: Preclinical Model Performance Metrics

Compared to the preclinical stage, the models' prodromal stage prediction abilities were not particulary higher but more reliable because of the larger dataset. While a population of 87 for preclinical is a high enough number it is important to mention that in the splitting of data this becomes a smaller number. The prodromal phase longitudinal model showed a strong ability to distinguish between people likely to go on to a later stage and those remained in the prodromal stage. The model's average C-index, which measures how well it predicts the progression of sickness, was 0.7389. In the prodromal stage, the estimated median survival time was 3.30 years, with a corresponding confidence range of 1.04 to 7.06 years. The model's concordance index of 0.70 further supported its capacity to classify people according to their likelihood of contracting the illness. Both the baseline and longitudinal models showed decent capabilities for differentiating across stages when compared to the models for the prodromal stage. The baseline model's average C-index of 0.7582 indicated a respectable capacity for discrimination, and the longitudinal model's average C-index of 0.7389 was comparable. In the prodromal stage of mild cognitive impairment, the predicted median surviving lifetime was 3.30 years, meaning that about half of the patients would go on to the next stage during this period. The baseline model's concordance value of 0.7322 showed a respectable level of prediction accuracy. Overall, the findings imply that the created disease progression models have the capacity to forecast the change from one stage of the disease to another. It is crucial to remember that these results are unique to the models and dataset utilized in the study. The generalizability and robustness of the models would need to be confirmed by additional analysis and validation on bigger and more varied datasets. The model's accuracy and performance should also be improved, notably in the preclinical stage where potential for improvement was found. A reason for the preclinical stage (N=87) being a little less reliable could be due to the fact that the population is smaller than prodromal stage (N=320). Thus gathering more data for both stages, or using a different dataset instead of ADNI with more information could result in even more precise progression models. Further investigation could

Model	**Average C-index**	**Survival Lifetime (95% CI)**	**C-index on test data**
Baseline	0.76	3.30 (1.01, 9.99)	0.73
Longitudinal	0.74	3.30 (1.04, 7.06)	0.70

Table 6.3: Prodromal Model Performance Metrics

also be possible by using other performance metrics such as AUC (Area under the curve) or look into the python libraries and find more such metrics. This could give even more information.

Chapter 7

Conclusions

This chapter gives the final conclusion after having done the research and also the possible future work which could further improve this thesis.

7.1 Conclusion

In order to better the lives of patients and their families, this thesis research has studied the devastating nature of Alzheimer's disease and the critical need for early detection and intervention. The assessment of preclinical and prodromal stages in patients with amyloid pathology was the main focus of this work, which significantly improved our comprehension of the course of the illness. This research has given important insights into the change from healthy cognition to cognitive impairment and dementia through the use of Artificial Intelligence techniques and the analysis of longitudinal data from the ADNI cohort. A more thorough comprehension of the root causes of Alzheimer's disease has been attained by developing biomarker predictors with high potential for forecasting disease progression and by looking at tau pathology and neuro-degeneration profiles. My understanding of Alzheimer's disease has increased because to the incorporation of numerous data sources and the use of cutting-edge analytical methods, which have also made it possible to create disease progression and prediction models. These discoveries may improve early detection and intervention tactics, enabling medical professionals to step in at a critical time and maybe halt the progression of the illness. Additionally, my thesis project has given me a priceless chance to advance my knowledge in data science and engineering, particularly in the context of machine learning. My awareness of the practical consequences of data-driven research in the healthcare field has grown as a result of applying these abilities to the study of intricate datasets related to Alzheimer's disease. My technical knowledge has increased as a result of this experience, but it has also strengthened my understanding of the value of multidisciplinary cooperation in addressing difficult medical problems. Additionally, this study path has given me a great awareness of how Alzheimer's disease affects people individually and families as a whole. My dedication to advancing Alzheimer's research and improving the lives of those afflicted has been strengthened by having direct experience with the disease's terrible impacts. In conclusion, the results of this thesis study have important implications for early detection and intervention efforts in addition to advancing our understanding of Alzheimer's disease. My skill set has been broadened and my comprehension of this crippling disease has been deepened by the

integration of data engineering, data science, and machine learning methodologies. I am appreciative of the chance to develop my expertise in these areas and to have made a contribution to the global fight against Alzheimer's disease.

7.2 Future work

7.2.1 Validation and Replication

The generalizability and robustness of the proposed models and biomarker predictors must be ensured through the validation and replication of the results in larger and more varied populations. The pooling of data from various sources would be facilitated by collaborative efforts with other research institutes and data sharing programs, allowing the assessment of the found biomarkers and prediction models across a larger variety of populations. Researchers may evaluate the consistency and reliability of the findings by incorporating datasets from several cohorts, such as AIBL or BioFinder, and analyzing their performance and prediction potential in various demographic and clinical scenarios. Additionally, having access to a variety of datasets would make it possible to investigate potential sources of heterogeneity, such as genetic variants or environmental influences, and how these can affect the development of diseases. Additionally, validation studies can examine the repeatability of the models and biomarkers in separate cohorts while using exacting statistical techniques to guarantee the accuracy of the findings. Overall, carrying out thorough validation and replication studies would expand the relevance of the findings to many groups, boost trust in them, and advance our understanding of the intricate dynamics of Alzheimer's disease.

7.2.2 Exploring Advanced Machine Learning Techniques

Exploring cutting-edge machine learning methods offers a fascinating chance to improve the precision and prognostication of the models created for this thesis. Although the COX model has been successful in capturing the dynamics of progression and survival analyses, additional research into different modeling strategies is necessary. This paper might get new insights and improve the performance in diagnosing illness development by using cutting-edge machine learning techniques like deep learning architectures and ensemble approaches. The models may automatically derive hierarchical representations from complicated and high-dimensional data by using deep learning algorithms, for example, which may be able to detect subtle patterns and relationships that conventional methods could have missed. Incorporating sophisticated optimization strategies and regularization approaches can also reduce overfitting and increase the generalizability of the model. Additionally, investigating other Python libraries and machine learning frameworks, like TensorFlow or PyTorch can give a wealth of tools and resources to successfully construct and enhance the models. The functionality of the models may be enhanced, improve prediction accuracy, and contribute to the most cutting-edge research in the area by adopting new breakthroughs in machine learning.

7.2.3 Expansion to Other Illnesses

Although the majority of this thesis has been devoted to the study of Alzheimer's disease, it is possible to gain important knowledge about the development and underlying processes of other neurodegenerative illnesses and associated disorders by broadening the scope of the research. For instance, including Parkinson's disease as research goals would make it possible to identify common biomarkers and create more thorough prediction models. Researchers might potentially detect overlaps in the pathophysiological mechanisms underlying these diseases and identify distinctive characteristics particular to each condition by examining the similarities and contrasts between these conditions. This larger viewpoint can help with the creation of individualized strategies for early detection, prognosis, and intervention while also enabling a more thorough knowledge of neurodegenerative disorders as a whole. Research on shared treatment targets or preventative measures can also be facilitated by the discovery of similar risk factors or comorbidities that contribute to the course of numerous illnesses. In the end, broadening the study's focus to incorporate more neurodegenerative conditions advances our understanding of these complicated diseases from a more comprehensive perspective and has the potential to have an impact on a wider spectrum of patients and their families.

Appendix A

Appendix

In this chapter we explore used theories.

A.1 Anova and Multicomparison

A.1.1 Introduction

ANOVA and multicomparison are crucial methods in the field of statistics. They assist in evaluating the differences in averages between multiple groups and draw inferences about the population as a whole.

A.1.2 Anova

To analyse the statistical differences between populations, it is mandatory to calculate the means of each group. If the means are not equivalent, a conclusion can be made that the groups differ from each other statistically. When there are just two groups, the t-test is widely used. When there are three or more groups this t-test cannot be used anymore and we have to switch to ANOVA (ANalysis Of VAriance), which is a method used to determine if there is a significant difference between three or more groups (by their means).

The principle of ANOVA as a test is relatively straightforward. Consider there are three groups namely Spanish (S), Belgian (B), and French (F), also assume that the height of 100 individuals has been measured in each group. Each individual is indicated by y(i), with y(i_S) denoting a Spanish person, y(i_B) a Belgian, and y(i_F) a French individual. A general average can be calculated from all the measured values, which is represented by μ. This average is the mean of all 300 individuals. A mean can be calculated for each group: $\overline{y(j)}$. Thus, there is an average for the height of the Spanish ($\overline{y_S}$), the Belgians ($\overline{y_B}$), and the French ($\overline{y_F}$). [12]

From the data the following three variances can be obtained:

1. **the variance between the groups** The variance arising from the difference between the group means and the general mean (y(j) compared to μ)

2. **The variance within the groups** The variance of each research unit relative to the group mean (all individuals in a group compared to the group mean, i.e., y(iS) compared to y(S), y(iB) compared to y(B), and y(iF) compared to y(F)).

3. **The total variance** The variance of each research unit compared to the general mean (all individuals compared to the general mean, i.e., y(i) compared to μ).

The complete variance (3) consists of two parts: the variance between groups (1) and the variance of individuals compared to the group mean (2). After calculating the percentage of the group variance accounts for the overall variance (= 1/3) and the percentage of the individual variance within groups accounts for the total variance (= 2/3), it is possible to determine how much of the total variance is caused by the differences in groups. Those that cannot be explained by group differences is called residual variance, unexplained variance or error.

The goal of ANOVA is to analyse if there are significant differences between given groups (like in the previous example of the height of the Spanish, Belgian and French subjects). If the differences between the groups are not significant, the respective means do not differ a lot. In the same way when there is in fact significant differences between the groups, a conclusion is made that there are dissimilarities between the groups.To verify whether the differences between these groups are meaningful, ANOVA is used. Nonetheless, The comparison is biased because of the fact that the dissimilarities between groups rely on a small number of groups, whilst the differences within the groups rely on many individuals. Thus, the variances should be categorized by the degrees of freedom first. Afterward, an F-value is computed and compared to an F-value in a reference table [13]. If the computed F-value is higher than the critical value in the table, it is considered that there is a statistically significant difference (according to Fisher's test theorem).

ANOVA using Python (Scipy)

Using the example of the heights from the three different nationalities. You put in all the measurements in an array, e.g. heightSpanish. And as the null hypothesis we assume that there is no statistical difference between the groups. With the following code you can perform one way ANOVA in python:

```
1. f_oneway(heightSpanish, heightBelgian, heightFrench)
```

After compiling this one line of code, an output is generated with an F score and a p-value. If the p-value is less than 0.05 it is concluded that there is a significant difference between the three groups and that the null hypothesis can be rejected. [26]

A.1.3 Multicomparison

Statistical inference is the process of generalizing about a population based on a sample. A common technique for statistical inference is hypothesis testing, where a null hypothesis is compared to an alternative hypothesis. The null hypothesis is considered to be true unless there is substantial evidence against it, while the alternative hypothesis is the reverse. The p-value is a metric that measures the evidence against the null hypothesis. It is the probability of observing a test statistic that is as extreme or more extreme than the one observed, assuming that the null hypothesis is not false. The p-value is then used to assess if the evidence against the null hypothesis is strong enough to reject it. A commonly accepted significance level is 0.05, meaning that if the p-value is less than 0.05, the null hypothesis is rejected and replaced by the alternative hypothesis. [31]

A.1.4 Statistical interference (p-value, null hypothesis)

Statistical inference is a way of drawing conclusions about a population based on a sample of data. By analyzing the data, you can figure out what the average or typical value is for that group of people, and how likely it is that different values fall within that range. The p-value is a measure of how likely it is that the results of a test statistic would have occurred by chance. If the p-value is small (less than 0.05), it means that there is a lot of evidence against the assumption that the results of the test statistic were due to chance. This suggests that the observed effect may be real and should be considered when making decisions about the null hypothesis.

The null hypothesis is a statement about the mean or average values of a group of people. It's always assumed until there's strong evidence to the contrary that the null hypothesis is true - in other words, that the population mean is equal to the given value. The alternative hypothesis is the opposite of the null hypothesis, and represents the idea that the population mean may be different from the given value. Statistical inference is used to test hypotheses about population parameters. Researchers use sample data to try to figure out what might be true about the population, and they use this information to decide if the data support or contradict the null hypothesis. [20]

Bonferroni Correction

When examining various groups, one-way ANOVA gives more information about the significant difference between these groups as said in previous sections. However it is not able to give insight about which parties are responsible for the difference. Additionally, there's more chance for false positives when multiple hypothesis test are conducted simultaneously. This is called a family-wise error rate which gauges the likelihood of commiting a false positive in any given hypothesis test. The Bonferroni correction improves and adjusts the significance level of multiple hypothesis tests to control the family-wise error rate (FWER). The FWER is the probability of a false positive among all the test. As more test are being undergone, this probability increases. The Bonferroni correction is used to counter this.

Suppose the means of k groups are being compared with a one-way ANOVA and then multiple pairwise comparisons are made between the groups. If a significance level of α for each test is used, the FWER will be greater than α. When the Bonferroni correction is being used the significance level (α) is divided by the number of comparisons being made (m). The significance level is now $\frac{\alpha}{m}$ for each test. E.g. when comparing the means of four groups while performing six pairwaise comparisons and using a significance level of 0.05, the correction would be $\frac{0.05}{6} = 0.0083$ for each test. If the p-value for a test is less than 0.0083, we reject the null hypothesis and conclude there is a significant difference. The Bonferroni correction is a conservative method, this means that it can increase the risk of false negatives because of the lowered significance level for each test making it harder to reject the null hypothesis. Despite this the correction is very useful when the number of test is small or when the conductor of a test wants it to be more conservative. [45]

Turkey Method

After performing an analysis of variance (ANOVA), Turkey's post hoc correction, also referred to as the Turkey-Kramer post hoc test or the Tukey's Honestly Significant Difference (HSD) test, is a statistical technique used in data analysis to identify significant differences between multiple groups or treatments. When doing several pairwise comparisons, this adjustment helps prevent an exaggerated Type I error rate. Researchers frequently compare the means of different groups while doing an ANOVA to see whether there are any statistically significant differences between them. A higher risk of making a Type I error—the rejection of a valid null hypothesis—can arise from doing many pairwise comparisons without any form of correction.

This problem is addressed by the Turkey-Kramer post hoc adjustment, which modifies the significance threshold for each pairwise comparison. Based on the total alpha level (often set at 0.05) and the number of groups or treatments being compared, the test determines a critical value. To ascertain if the observed differences between groups are statistically significant, the sample size and data variability are taken into consideration. When the null hypothesis in an ANOVA is rejected, suggesting that there are substantial differences between the groups, the Turkey's post hoc adjustment is frequently applied. It assists in determining which particular groups differ greatly from one another. The test assesses if the observed differences are statistically significant by comparing the difference between the means of each pair of groups with the critical value. For each pair of group means, the correction technique entails computing a studentized range statistic, also referred to as the q statistic. When standard error and sample size are taken into account, this statistic calculates the actual difference between the means. The crucial value produced from the distribution of the range statistic is then used to compare the q statistic against.

The difference between the means of the two groups is statistically significant at the given alpha level if the absolute value of the q statistic exceeds the critical value. The test compares the difference between the means of each pair of groups with the critical value to determine if the observed differences are statistically significant. The correction method involves calculating a studentized range statistic, commonly known as the q statistic, for each pair of group means. This statistic determines the true difference between the means after accounting for standard error and sample size. The q statistic is then measured against the critical value derived from the range statistic's distribution. [42] If the absolute value of the q statistic is greater than the threshold value, the difference between the means of the two groups is statistically significant at the specified alpha level. It is crucial to remember that the Turkey-Kramer test makes the assumptions that the data have a normal distribution and that the variances between the groups are equal. The validity of the test findings may be impacted if certain assumptions are broken. Larger samples are more likely to produce statistically significant findings since the test is also size sensitive. In conclusion, the Turkey's post hoc correction, a useful statistical technique that aids in the discovery of significant differences between several groups after an ANOVA, is used by researchers. It allows researchers to perform trustworthy comparisons and derive insightful inferences from their data by limiting the Type I error rate. [56]

Bibliography

[1] ADNI (Alzheimer's Disease Neuroimaging Initiative). http://adni.loni.usc.edu/.

[2] ADNI Logo. http://adni.loni.usc.edu/. Image retrieved from ADNI website.

[3] Jupyter Notebook. https://jupyter.org/.

[4] Jupyter Notebook Logo. https://github.com/jupyter/jupyter.github.io. Image retrieved from Jupyter Project.

[5] Magerit-3. https://www.cesvima.upm.es/infrastructure/magerit.

[6] Magerit-3. https://www.xataka.com/otros/magerit-el-nuevo-supercomputador-de-la-upm.

[7] MATLAB Logo. https://www.mathworks.com/company/newsroom/media-resources.html. Image retrieved from MathWorks.

[8] PyCharm Logo. https://www.jetbrains.com/pycharm/. Image retrieved from JetBrains.

[9] Python Programming Language. https://www.python.org/.

[10] R Logo. https://www.r-project.org/logo/. Image retrieved from R Project.

[11] The R Project for Statistical Computing. https://www.r-project.org/.

[12] Anova. https://hulpbijonderzoek.nl/online-woordenboek/begrippen/anova/, 13th April 2023.

[13] Ravinder K. Agarwal. *Analysis of Variance: One-Way ANOVA*. Springer, New Delhi, 2018.

[14] Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottan, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibeau-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot, and Stanley Durrleman. Impact of method characteristics, 2021.

[15] Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottan, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibeau-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot, and Stanley Durrleman. Impact of method characteristics, taking into account the interaction between the model type and the use of imaging features, 2021.

[16] Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottan, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibeau-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot, and Stanley Durrleman. Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review. *NeuroImage: Clinical*, 31:102744, 2021.

[17] Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottan, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibeau-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot, and Stanley Durrleman. Recent trends, 2021.

[18] Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottan, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibeau-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot, and Stanley Durrleman. Recent trends (2), 2021.

[19] Manon Ansart, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Bottan, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibeau-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot, and Stanley Durrleman. Relationship between the auc (area under the roc curve) and the number of indi- viduals., 2021.

[20] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Press, Belmont, CA, 2nd edition, 2002.

[21] Michael C Donohue, Hélène Jacqmin-Gadda, Mélanie Le Goff, Ronald G Thomas, Rema Ramana, Anthony C Gamst, Laurel A Beckett, Clifford R Jack Jr, Michael W Weiner, Jean-Francois Dartigues, and Paul S Aisen. Adni amyloid+ subjects, 2014.

[22] Michael C Donohue, Hélène Jacqmin-Gadda, Mélanie Le Goff, Ronald G Thomas, Rema Ramana, Anthony C Gamst, Laurel A Beckett, Clifford R Jack Jr, Michael W Weiner, Jean-Francois Dartigues, and Paul S Aisen. Adni panel of biomarkers and assessments, 2014.

[23] Michael C Donohue, Hélène Jacqmin-Gadda, Mélanie Le Goff, Ronald G Thomas, Rema Ramana, Anthony C Gamst, Laurel A Beckett, Clifford R Jack Jr, Michael W Weiner, Jean-Francois Dartigues, and Paul S Aisen. Counts of subjects and observations by baseline diagnosis, 2014.

[24] Michael C Donohue, Hélène Jacqmin-Gadda, Mélanie Le Goff, Ronald G Thomas, Rema Ramana, Anthony C Gamst, Laurel A Beckett, Clifford R Jack Jr, Michael W Weiner, Jean-Francois Dartigues, and Paul S Aisen. Estimating long-term multivariate progression from short-term data. 10(5):S400–S410, 2014.

[25] Cleusa P Ferri, K S Jacob, Dementia, and Working Group. Global prevalence of dementia: a delphi consensus study. *The Lancet*, 396(10251):2135–2143, 2020.

[26] GeeksforGeeks. How to perform a one-way anova in python. `https://www.geeksforgeeks.org/how-to-perform-a-one-way-anova-in-python/`, 13th April 2023.

[27] Oskar Hansson, John Seibyl, Erik Stomrud, Henrik Zetterberg, John Q. Trojanowski, Tobias Bittner, Valeria Lifke, Veronika Corradini, Udo Eichenlaub, Richard Batrla, Katharina Buck, Katharina Zink, Christina Rabe, Kaj Blennow, and Leslie M. Shaw. Csf biomarkers of alzheimer's disease concord with amyloid-b pet and predict clinical progression: A study of fully automated immunoassays in biofinder and adni cohorts. *Alzheimer's & Dementia*, 17(1):29–44, 2021.

[28] Oskar Hansson, John Seibyl, Erik Stomrud, Henrik Zetterberg, John Q. Trojanowski, Tobias Bittner, Valeria Lifke, Veronika Corradini, Udo Eichenlaub, Richard Batrla, Katharina Buck, Katharina Zink, Christina Rabe, Kaj Blennow, and Leslie M. Shaw. Distribution of the csf biomarkers colored by pet visual read classification., 2021.

[29] Oskar Hansson, John Seibyl, Erik Stomrud, Henrik Zetterberg, John Q. Trojanowski, Tobias Bittner, Valeria Lifke, Veronika Corradini, Udo Eichenlaub, Richard Batrla, Katharina Buck, Katharina Zink, Christina Rabe, Kaj Blennow, and Leslie M. Shaw. Performance of csf biomarker cutoffs versus visual amyloid-β pet in biofinder and adni, 2021.

[30] Oskar Hansson, John Seibyl, Erik Stomrud, Henrik Zetterberg, John Q. Trojanowski, Tobias Bittner, Valeria Lifke, Veronika Corradini, Udo Eichenlaub, Richard Batrla, Katharina Buck, Katharina Zink, Christina Rabe, Kaj Blennow, and Leslie M. Shaw. Scatterplots of csf biomarkers versus suvrs in biofinder (a–c) and adni (d–f), 2021.

[31] Yosef Hochberg and Ajit C. Tamhane. *Multiple Comparison Procedures*. John Wiley & Sons, New York, NY, 1995.

[32] Jack. At(n) classification.

[33] Jack. Syndromal cognitive stage.

[34] Clifford R Jr Jack, David A Bennett, Kaj Blennow, Maria C Carrillo, Billy Dunn, Samantha Budd Haeberlein, David M Holtzman, William Jagust, Frank Jessen, Jason Karlawish, Enchi Liu, Jose Luis Molinuevo, Thomas Montine, Creighton Phelps, Katherine P Rankin, Christopher C Rowe, Philip Scheltens, Eric Siemers, Heather M Snyder, and Reisa Sperling. Nia-aa research framework: Toward a biological definition of alzheimer's disease. *Alzheimer's dementia*, 14(4):535–562, 2018.

[35] JetBrains. PyCharm - Python IDE. `https://www.jetbrains.com/pycharm/`.

[36] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[37] David G Kleinbaum and Mitchel Klein. *Survival analysis: A self-learning text*. Springer Science & Business Media, 2005.

[38] Igor O Korolev, Laura L Symonds, Andrea C Bozoki, and Alzheimer's Disease Neuroimaging Initiative. Comparison of models for predicting mci-to-ad progression, 2020.

[39] Igor O Korolev, Laura L Symonds, Andrea C Bozoki, and Alzheimer's Disease Neuroimaging Initiative. Cross-validated performance estimates for single-source and multi-source models, 2020.

[40] Igor O Korolev, Laura L Symonds, Andrea C Bozoki, and Alzheimer's Disease Neuroimaging Initiative. Predicting progression from mild cognitive impairment to alzheimer's dementia using clinical, mri, and plasma biomarkers via probabilistic pattern classification. *Journal of Alzheimer's Disease*, 74(4):1205–1218, 2020.

[41] Igor O Korolev, Laura L Symonds, Andrea C Bozoki, and Alzheimer's Disease Neuroimaging Initiative. Subject characteristics at baseline, 2020.

[42] Cyrus Y. Kramer. *Extension of multiple range tests to group means with unequal numbers of replications*, volume 12. Biometrics, 1956.

[43] MathWorks. MATLAB. `https://www.mathworks.com/products/matlab.html`.

[44] Benjamin B Miller and Eric Karran. Ethical considerations in alzheimer's disease: challenges and recommendations. *Trends in Neurosciences*, 41(12):1058–1069, 2018.

[45] OpenStax. *Introductory Statistics*. OpenStax, 1.0 edition, 2017.

[46] Nicole Palmour, Amadou Ndao, Augustine Choko, Anne-Andrée Baril, and Éric Racine. Ethical considerations in alzheimer's disease research: Challenges and recommendations. *Canadian Journal of Bioethics*, 2(3):1–16, 2019.

[47] Platero and Tobar. Categorical predictive and disease progression modeling in the early stage of alzheimer's disease.

[48] Platero and Tobar. Cross-validation procedure for predictive model development and evaluation.

[49] Platero and Tobar. oefficients associated with the markers (in z-score values).

[50] Platero and Tobar. Time course of ptau/aβ(1–42) ratio in patients with mci in the adni cohort over 2 years. ls-means with standard errors by biomarker group.

[51] Carlos Platero and M. Carmen Tobar. Longitudinal survival analysis and two-group comparison for predicting the progression of mild cognitive impairment to alzheimer's disease. *Journal of Neuroscience Methods*, 341:108698, 2020.

[52] Carlos Platero and M. Carmen Tobar. The most significant p-values when comparing values between clinical groups at baseline (c1) and over time (c2)., 2020.

[53] Carlos Platero and M. Carmen Tobar. Scores for predicting mci-to-ad conversion using multisource biomarkers., 2020.

[54] Carlos Platero and M. Carmen Tobar. Scores for predicting mci-to-ad conversion using only mri-based biomarkers., 2020.

[55] Carlos Platero and M. Carmen Tobar. The smoothed longitudinal trajectories of some of the best biomarkers under the univariate analysis, 2020.

[56] John W. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 3(1):99–114, 1953.

[57] Lisa Vermunt, Sietske A. M. Sikkes, Ardo van den Hout, Ron Handels, Isabelle Bos, Wiesje M. van der Flier, Silke Kern, Pierre-Jean Ousset, Paul Maruff, Ingmar Skoog, Frans R. J. Verhey, Yvonne Freund-Levi, Magda Tsolaki, Asa K. Wallin, Marcel Olde Rikkert, Hilkka Soininen, Luisa Spira, Henrik Zetterberg, Kaj Blennow, Philip Scheltens, Graciela Muniz-Terrera, and Pieter Jelle Visser. Baseline characteristics according to diagnosis, 2021.

[58] Lisa Vermunt, Sietske A. M. Sikkes, Ardo van den Hout, Ron Handels, Isabelle Bos, Wiesje M. van der Flier, Silke Kern, Pierre-Jean Ousset, Paul Maruff, Ingmar Skoog, Frans R. J. Verhey, Yvonne Freund-Levi, Magda Tsolaki, Asa K. Wallin, Marcel Olde Rikkert, Hilkka Soininen, Luisa Spira, Henrik Zetterberg, Kaj Blennow, Philip Scheltens, Graciela Muniz-Terrera, and Pieter Jelle Visser. Duration of preclinical, prodromal, and dementia stages of alzheimer's disease in relation to age, sex, and apoe genotype. *Alzheimer's & Dementia*, 17(2):261–271, 2021.

[59] Lisa Vermunt, Sietske A. M. Sikkes, Ardo van den Hout, Ron Handels, Isabelle Bos, Wiesje M. van der Flier, Silke Kern, Pierre-Jean Ousset, Paul Maruff, Ingmar Skoog, Frans R. J. Verhey, Yvonne Freund-Levi, Magda Tsolaki, Asa K. Wallin, Marcel Olde Rikkert, Hilkka Soininen, Luisa Spira, Henrik Zetterberg, Kaj Blennow, Philip Scheltens, Graciela Muniz-Terrera, and Pieter Jelle Visser. Results of vermunt et. al, 2021.

[60] Anders Wimo, Melanie Gaudig, Carla Schölzel-Dorenbos, Steffi Riedel-Heller, Rik Vandenberghe, Gunhild Weyer, Bengt Winblad, Ake T Sölund, Richard Dodel, Uwe Zilkens, et al. The economic impact of dementia in europe in 2015–cost estimates from the eurocode project. *The Lancet Neurology*, 16(6):766–771, 2017.